Artificial Curiosity for Autonomous Space Exploration

Vincent Graziano, Tobias Glasmachers,
Tom Schaul, Leo Pape,
Giuseppe Cuccu, Jürgen Leitner,
and Jürgen Schmidhuber

Abstract

Curiosity is an essential driving force for science as well as technology, and has led mankind to explore its surroundings, all the way to our current understanding of the universe. Space science and exploration is at the pinnacle of each of these developments, in that it requires the most advanced technology, explores our world and outer space, and constantly pushes the frontier of scientific knowledge. Manned space missions carry disproportionate costs and risks, so it is only natural for the field to strive for autonomous exploration. While recent innovations in engineering, robotics and AI provide solutions to many sub-problems of autonomous exploration, insufficient emphasis has been on the higher level question of autonomously deciding what to explore. *Artificial curiosity*, the subject of this paper, precisely addresses this issue. We will introduce formal notions of "interestingness" based on the concepts of (1) compression progress through discovery of novel regularities in the observations, and (2) coherence progress through selection of data that "fits" the already known data in a compression-based way. Further, we discuss how to construct a system that exhibits curiosity driven by the interestingness of certain types of novel observations, with the mission to curiously go where no probe has gone before.

1 Introduction

Technology, and science in general, has progressed to the point that we can send intricate machines which are computationally powerful into space. So advanced are these machines that their potential easily exceeds what can be accomplished by manual control from Earth: the bandwidth of the sensors, such as cameras, usually exceeds the bandwidth of the communication channel between space probe and mission control, and more importantly, the enormous time latency involved in the exchange of signals imposes strict limitations on the efficiency of manual control. The efficiency of these probes could be greatly enhanced if they were able to carry-out missions and explore autonomously. For efficient autonomous exploration a probe needs to be as like a scientist as possible. In this paper we discuss the concept of *artificial curiosity* [12, 15] and how it can be used to best approximate the decisions a scientist might make, guided by his own knowledge, understanding, and curiosity.

Ideally, the device controller should make informed decisions on which information to send back to Earth, and what to do next in order to optimally achieve its mission goals. However, building such a controller solely based on assumptions and information available years before a mission begins does not really solve the problem. The very nature of an exploratory mission is to encounter novel, and unexpected information. One cannot provide *a priori* behaviors to efficiently study unpredicted phenomena. A more functional design might include on-board control software capable of autonomously classifying a piece of information such as an image or a situation as interesting or not worthwhile to investigate or even transmit back to Earth, based on the goals of the spacecraft. The probe could also then send interesting information to Earth with priority over non-interesting information. More importantly, rather than to wait for mission control to analyze the information and send back an appropriate action plan, our space probe could actively explore the interesting phenomenon until it has gained some understanding thereof and no longer classifies it as interesting. Different phenomena would then become promising sources of novel information and attract the interest of our probe.

However, information collected this way serves another purpose beyond being reported to Earth: it may facilitate decision making in the future, enabling the agent to bootstrap detection and understanding of complex phenomena, based on simpler ones learned earlier in the mission. This automatic adaptation to previously unknown phenomena further increases the probe's autonomy from mission control. It enables the probe to respond more appropriately to circumstances not foreseen at design time than any catch-all algorithm might.

Intelligent autonomous control in active agents such as space probes is addressed within the field of artificial intelligence. A control algorithm or 'agent', such as a so-called reinforcement learner [5, 23], abstracts a spacecraft into two types of components: sensors that provide observations of the environment, and actuators that effect the surrounding environment, including the agent itself. The control algorithm bridges the sensory and motor systems. In reinforcement learning language, we talk about an agent or controller, that learns a policy which determines the actions taken by the agent, given the history of observations and actions.

For every interaction with the environment, be it exploratory or otherwise, the algorithm updates the policy based on success or failure of the action sequence attempted. The measure of success or failure is usually encoded into a single *reward* signal [23]. This way the agent learns from experience and autonomously improves its policy over time. However, the big open challenge in this learning paradigm is how to decide where to look and what to try next in order to maximize learning progress. Humans seem to make such decisions with relative ease, driven by their internal curiosity. The idea behind artificial curiosity is to transfer this internal human drive to reinforcement learning, making autonomous exploration practical. This amounts to providing the learning agent with an automatically available internal curiosity reward signal.

In this article we describe a route for transforming a probe into a more autonomous agent. We will achieve this goal in two conceptual steps, which are both related to the design of an internal feedback signal guiding curiosity. First we provide the probe with a way to measure the interestingness of its sensory inputs, which allows it to passively judge the observations it makes. Then we come to the active part, the crux of autonomy, namely how to come up with action plans or behaviors that lead to exciting new observations, which is known as *artificial curiosity* [11, 13]. We close the article with a discussion on human curiosity as a driving force of scientific research.

Let us begin by introducing a number of concepts prerequisite to the notion of artificial curiosity.

2 What Is Interesting?

Although we are aiming for an autonomous decision maker, let us first restrict ourselves to an agent which cannot take any actions, e.g., a probe which passively monitors its environment. Here, the design goal for an autonomous agent reduces to detecting interesting pieces of information among the vast stream of incoming observations. To thoroughly address this problem we need a solid definition of interestingness, as well as practical algorithms for classifying information as such. However, the notion of interestingness, although intuitively clear, got formalized only relatively recently, using the concept of *learning progress* [11], in particular *compression progress* [18].

Let us look at an example. Assume a probe (a rover) equipped with sensors to continuously monitor its surrounding. Although we are ultimately interested in making the rover explore its surroundings autonomously let us for the moment focus only on specific observations: the simple thermal sensor records, a single number at each time step, capturing some information about its environment. Let us assume that the sensor records its readings as plain text strings, consisting of the time of the reading, taking a measurement each minute, and the temperature in Kelvin, with an accuracy of $1/10$ Kelvin. A typical record might read:

12/25/2030 14:11 — 288.6 K,

where the first entry indicates the date and time at mission control when the reading was taken. This is one possible encoding of the information, in a both human and machine readable format. Of course, this is an arbitrary choice, and there are many ways to encode the same information, some more compact than others. For example, including the start date and time elapsed in every record makes each record a complete piece of information, but it is highly redundant within the stream of records. We can save storage space by noting the start date only once. Knowing that the temperature is recorded once a minute means that we do not even need a time stamp at all; the time of a measurement can be deduced from the time the series of measurements started and the index of the record in the list. Moving to a smarter encoding that saves storage space is known as *compression* [7]. Saving storage may or may not be a big deal for our space probe, but it surely is when sending the information through a limited bandwidth channel. Moreover, we will see in the following that the ability to compress information can have far-reaching implications for our explorer.

Given that using shorter encodings is beneficial, is there a better way to compress our data? The first step in compression is to look for *patterns* in the data. A pattern can be defined as something that, on any level of abstraction, repeats itself. For example, fractals possess self-similarity and repeat their pattern at every scale, the decimal expansion of π is the result of the repeated application of the same numerical procedure. Once a stream of data is analyzed, and repetitive trends are found, they can be exploited for the construction of a *model* of the pattern. When analyzing our temperature data we will find a profile which, more or less, repeats with each rotation of the planet. The values will rise in the morning and drop in the evening. A compressor can store this pattern once and then use it as a model to compress future observations. It makes use of this model by storing only the differences of the measurements from the model. These representations of the temperatures will, on average, be smaller than the actual values, thereby shortening the length of the encoding.

Typical (loss-less) general purpose compression algorithms exploit different types of patterns. One strategy is to build up a dictionary of frequent patches of information, patterns, which can be referenced with short codes. In contrast, statistical encoding relies on a probability model of which information is expected to occur next, such that the most probable next temperature can be encoded with a very short code [3], while for example rare jumps in temperature will have longer codes. These two strategies are often combined, such as in the famous Lempel-Ziv-Welch algorithm [24], which is the basis of the `gzip` tool, an often-used program for data compression on computers.

A model that fits the data well allows a compressor to code the information compactly. This means that the optimal model is problem-dependent. In the extreme case we arrive at the notion of Kolmogorov complexity [6, 7, 21]. The Kolmogorov complexity of a sequence is defined as the length of the shortest program (in a universal programming language) encoding this information. This formal approach to compression turns out not to be realizable in practice. Instead, one has to fall back to approximations and heuristics, such as dictionaries and statistical prediction, and be content with compressors that at best only approximate the ideal compressor and the shortest possible encoding of the data. However, even such imperfect compressors can be extremely powerful. For example, the models built by the human brain, which include tremendously useful concepts such as hierarchies (ranging from abstract to concrete notions), are all of the latter type, and can be realized as instances of compression. In addition, compression is intimately related to the concept of Occam's razor: the principle of preferring models with fewer assumptions (among the models that explain the data) translates to choosing the shortest program that compresses the data [1, 21].

We want to emphasize that the compressor has *learned* something about the environment by studying it, discovering patterns, and then constructing a model from these patterns. Building a model which explains some aspect of the environment, in the case of our probe the typical daily behavior, not only allows us to compress observations, but also represents what we have learned about the environment, and allows us to make *predictions* about future events [4, 21]. Applying the model to the (past) events used to construct the model itself allows the compressor to encode these events with short codes, resulting in data compression. We use the term 'prediction' for information obtained from the model, even when applied to data from the agent's history.

As discussed above, the notions of pattern, model, prediction and compression are intimately related. Learning is then the process of finding patterns in data and incorporating them into a model, which again allows the compressor to encode with shorter codes, and thus to compress and predict better. Although predictability and compressibility are not *quite* the same [9], we can use a measure of compression to express the quality of predictions a model can make about information in the environment. We will see that this notion also leads to a straight-forward definition of interestingness.

The process of learning or training in this passive scenario of monitoring the world and processing data is understood as *learning to predict* (in contrast to learning to act well). This is the same as learning to compress. The goal of learning is thus equivalent to finding shorter and shorter codes for a given stream of data. However, as noted above, there is a lower bound on the length of an encoding that a compressor can find. Thus, the learning progress, measured by the decrease in length of the compressed representation, will essentially vanish as the length of the code approaches the length that an ideal compressor would produce. If the model is suitable for the problem at hand, this means that it will have learned to predict all the data arbitrarily well, leaving nothing else to learn. So what makes data or information *interesting*? Continuing the line of thought above, the extent to which new information is interesting is related to how much the model of the environment stands to improve by observing it. This concept is caught compactly by the notion of *compression progress*. Using the connection between prediction and compression, we now phrase the statement as such: information is interesting if it allows us to more succinctly code the observations we have made in our environment.

The drive to actively seek out data which allows for the learning of more expressive and compact models over time (or in other words, the drive to compress data with shorter programs) is known as *curiosity* [15]. The goal of curiosity is to maximize the pace of learning patterns and regularities. A rover can use this drive to learn new rules to govern its exploration strategy.

3 Compression Progress

Now that we have introduced and justified the utility and role of compression progress, we show how to formalize that notion. Broadly speaking, compression progress is a measure of how much the ability to compress the history of observations improves by either learning new patterns (from the history) or by making new observations.

We will discuss several aspects of this: (1) an adaptive compressor can learn previously unknown patterns and regularities by revisiting the history (by way of a change in the compressor), (2) new observations might yield additional data obeying unknown but learnable laws, (3) new observations might also increase the internal *coherence* of the history, leading to a better compressibility of the augmented history, even without a change in the compressor (by way of a change in the data).

3.1 Coherence progress

To simplify matters, this paper will mostly focus on the last aspect, *coherence progress* [10]. We assume a fixed compressor; for example, say, the `gzip` tool. The progress of such an algorithm may change as new observations arrive. A new observation can be compressed in the context of the history, and the overall compressibility of the history changes accordingly. Hence, the progress is directly attributed to the current observation, and results in a (partial) measure of its interestingness.

To formalize this, we first introduce the auxiliary concept of *compression similarity*

$$S(a, b) = L(a) + L(b) - L(a + b) \ ,$$

which is a measure of how closely two sequences a and b are related to each other. Here, $L(\cdot)$ is the length of a compressed sequence when using a fixed compressor. In other words, $S(\cdot, \cdot)$ is a measure of how many bits can be saved by compressing two sequences together, as opposed to compressing them separately. We can generalize this to the notion of *compression coherence* of a single sequence:

$$C(h) = \frac{1}{n-1} \sum_{i=1}^{n-1} S(h_{1:i}, h_{i+1:n}).$$

In words, coherence is the average compression similarity between two partitions of the history, cut at index i. Coherence progress can then be defined as the increase in compression coherence, when incorporating a new observation into the history, measured in bits:

$$\begin{aligned} P_h(o_{n+1}) =& C(h_{1:n} + o_{n+1}) - C(h_{1:n}) \\ =& C(h_{1:n+1}) - C(h_{1:n}), \end{aligned}$$

3.2 Compression Progress

Now let us turn to the general case of learning compressors, the focus of most previous work on interestingness [11, 22, 14, 15, 16, 19, 18, 17, 20]. In contrast to a fixed tool such as `zip`, an adaptive compressor is able to make compression progress simply by re-visiting the history, that is, without the need of additional observations [15, 20] (although additional observations may make it *easier* to achieve compression progress). This progress is achieved by a change in the compression strategy, which can be interpreted as a gain in understanding of the history. For example, a search process may find a new rule that allows the compressor to better predict forthcoming observations, thus encoding them from then on with shorter codes. Or an adaptive architecture, such as a recurrent artificial neural network, might adapt its synaptic weights to better reflect some aspect of the dynamics of the environment, again resulting in an even more powerful anticipation of observations.

Processes that lead to compression progress without additional observations are coupled with improvements of existing models, or with the emergence of completely new models. As discussed earlier, this added predictive power amounts to better understanding the sequence of observations, and thus the environment. This way, an adaptive compressor can adapt to its environment, building increasingly more complex rules on top of the pre-existing ones.

Let us treat this situation of an adaptive compressor in the terms of the length function L and the history $h_{1:n}$ introduced above. By re-visiting and re-compressing the history, an adaptive compressor can change its encoding, and thus the length function itself. It is straight-forward that the progress of a change

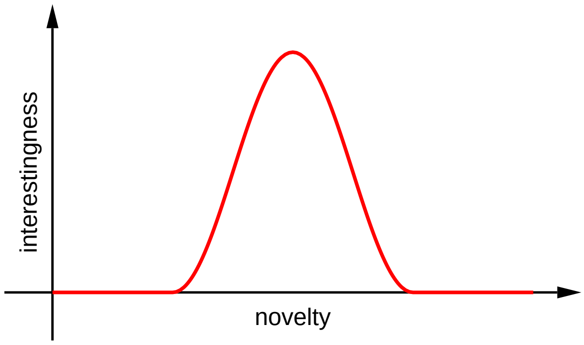

Figure 1: A Wundt Curve. Measures of interestingness date back to Wundt [25]. Wundt's curve shows the interestingness of an observation as a function of novelty. Novelty, unlike complexity, depends on the relationship between the information and the person observing it. Trivial patterns quickly lose their novelty, while noise is always novel. As learning proceeds the complexity of the most interesting patterns increases. From the artificial curiosity point of view, novelty can be considered inversely proportional to compressibility: noise has a low compressibility, and trivial or simple patterns have a high compressibility.

of encoding, resulting in a new length function $\tilde{L}$, should be measured by

$$L(h_{1:n}) - \tilde{L}(h_{1:n}) \ .$$

Even more general measures of learning progress also take into account the time needed for compressing and decompressing the data [15]. Because of the nature of restructuring its compression strategy by means of detecting additional patterns and gaining improved understanding of the history, this type of progress cannot be attributed solely to a particular observation. Rather, it is also to be attributed to the restructuring process itself. Although learning a better compressor may be costly in terms of computational resources such as time and memory, this process too can have a measure of interestingness attributed with it, one that is consistent with the measure used with new observations.

3.3 Relation of Interestingness to Coherence Progress

Setting aside the interestingness associated with learning a better compressor, we now return to the compression progress-independent interestingness associated to particular observations. Our measure of coherence progress already captures a number of desirable properties of interestingness.

For the purposes of the present section, a new observation is uninteresting when we cannot make coherence progress from it. This may happen for completely different reasons: The observation may be easily compressible, either because the observation is trivially compressible by itself (e.g., a long string of zeros, $L(o_{n+1}) \approx 0$), or because its information is redundantly present in the observation history already (the sunrise temperature increase, after having observed hundreds of sunrises before), if the underlying pattern (here, periodicity) has been discovered already, i.e., $L(h_{1:n+1}) \approx L(h_{1:n})$. In both these cases, $P_h(o_{n+1})$ will be low.

On the other hand, patterns may be so complex that we cannot find predictive models. This may happen either because we lack the necessary prerequisites (such as basic skills or knowledge) for discovering the patterns, or because the observation is actually random; in both cases we have $L(h_{1:n+1}) \approx L(h_{1:n}) + L(o_{n+1})$, and thus $P_h(o_{n+1})$ will be low. In principle it is therefore difficult to tell a random phenomenon from a pattern we fail to catch. However, given enough observations with an underlying regularity, and a compressor that is able to find this regularity, the history of observations can be stored in a much shorter form; for example, a child will find a course on advanced statistics completely boring, while a student with the necessary prerequisites may be fascinated. See Figure 1 for an informal illustration of how interestingness relates to the compressibility of observations.

It is important to note that data are not inherently interesting. What is currently interesting depends on context, namely, what we already know. For $P_h(o_{n+1})$ to be large, the new coherence $C(h_{1:n} + o_{n+1})$ needs to exceed the previous coherence $C(h_{1:n})$, which again means that new observations need to support the discovery of patterns in our experience for which we did not have sufficient evidence before. In order to keep things interesting we may profit from continually discovering new patterns. Consequently, as we learn more, that is as our predictive model becomes stronger, we have to turn to environments where pattern discovery was originally too difficult. For example, we finally enroll in the advanced statistics course. This happens for two reasons: (i) contexts in which pattern detection was extraordinarily difficult appear relatively more interesting because we already figured out the simpler patterns and (ii) more importantly, having discovered a base of rules by first exploring simpler environments allows us to extend these rules to more complex patterns in more complex environments. A result of learning is that we can learn things we previously were unable to. Figure 2 illustrates this point by showing how the values of coherence progress evolve as observations (from a given class) are accumulated. The coherence progress is measured for three different

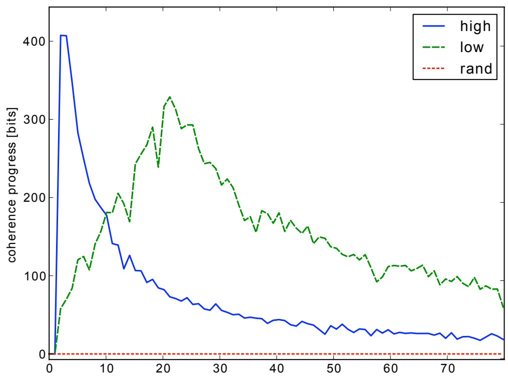

Figure 2: Illustration of the qualitative effect of aspects of interestingness, measured in terms of coherence progress $P_h(o_{n+1})$ achieved by a new observation, when added to the current history $h_{1:n}$ (where n is on the horizontal axis). The time profile of coherence progress evolves differently, depending on the amount of overlap between individual observations. The three classes from which the observations are drawn are the following. The plot in blue (solid line) shows observations with *high overlap*, where every observation shares 80% of its information with every other one (individual stones, say). For them coherence progress is high initially, but quickly decays, as the underlying patterns are easy to spot and then compress – observations from $n = 15$ onwards convey but few new insights. The plot in green (broken line) shows observations with *low overlap* (approximately 10%, different planets, say), in which case coherence progress only kicks in after the history has accumulated sufficient observations (around 20 here) for each new one to increase the coherence significantly (e.g., when different planet categories start to emerge). This class has more total information, and a more complex underlying structure, which is the reason why coherence progress stays comparatively high for a long time. In red (dotted), we show the plot for a sequence of *random* observations: the coherence progress is zero throughout, as expected.

classes, each with different degrees of shared information between the observations. Among other things, it also shows that there can be a trade-off between short-term and long-term progress.

Given that our formal definition of coherence progress provides a quantitative measure of certain aspects of interestingness, correctly capturing certain key qualitative aspects (noise is never interesting, redundant information quickly becomes uninteresting), that is simple to compute and understand, we will use it in the next section as a feedback signal that can inform an autonomous agent on where to explore next. In this way we are able to base a variant of artificial curiosity, the informed drive to explore aspects of the environment that maximize coherence progress, on a qualitative measure.

4 Choosing Actions

An agent equipped with intrinsic motivation, such as artificial curiosity, is able to control its behavior and steer itself autonomously towards places and phenomena that are *quantitatively interesting*. Such an agent, clearly, could play an important role in space exploration. We now discuss how to realize such an agent, making an autonomous robot more of a curious scientist. We want the agent to chose an action that results in interesting observations. How can we create such a curious agent?

Firstly, we need to extend our agent, which so far only does passive monitoring. Obviously, it is not enough for an agent to simply observe changes in its environment if it is to seek out interesting data. It must be able to perform different actions (e.g., move a sensor or its entire self), and more to the point, it must choose actions which actually lead to the observation of interesting data.

Without a good model of the world that agent can not know in advance which actions will most probably result in interesting observations. At the same time, without exploration the agent does not have access to observations from which it could build a sufficiently detailed model of its environment. Artificial intelligence research has resolved this dilemma by using, e.g., reinforcement learning algorithms [23]. Such algorithms are designed to bootstrap both action policy and world model at the same time. Typically, such algorithms are applied to autonomously achieve a pre-defined goal, encoded in a reward signal. Such a goal-related reward signal in known as *external reward* in the literature. With new observations and reward signals becoming available, reinforcement learning algorithms change their policy to make actions that have been rewarding in the past more likely in the future. This feedback, over time, shapes the behavior of the system, enabling it to evolve specialized strategies that achieve arbitrary pre-specified goals.

The decisive trick that turns an agent into a curious explorer is to provide the very same class of algorithms with a different type of feedback signal, namely with so-called *internal* or *curiosity reward* [12, 20]. This is the most direct way to reflect the drive for finding interesting observations in a control algorithm. Let us have a closer look at its different components in the following.

The agent needs to have a *model* of the world that represents what has been learned about the environment so far. Again, the role of this model is twofold: it allows for a compression of the history of its observations, and more relevant here, it will allow the agent to make predictions about future events. The agent also needs *a way to update* the internal world model as new observations are made. Most of the time it is sufficient to refine the existing model, but sometimes it may be necessary to find a completely fresh model, one that integrates newly discovered phenomena with the previous model and possibly resolves conflicts between old assumptions and new data. Such learning and updating of models based on data is a prototypical task for machine learning algorithms, and can be implemented for example with artificial neural networks.

In order to choose better actions the robot needs *feedback* on how well it is performing. In the reinforcement-learning paradigm, such feedback is encoded into a reward signal. Based on the observations that follow an action, we can either increase or decrease the likelihood of the action in similar future situations.

In the case of a curious explorer this goal can be cast as a drive to improve the internal world model, for which the curiosity reward is received. From previous sections we know that such improvements can be measured in terms of compression progress, hence, the curiosity reward feedback is the progress made by the system. Note that this curiosity feedback signal cannot be formulated as a function of the state of the agent and its environment. Instead, it depends on the agent's internal state, particularly on the predictive power of its current world model.

Let's now extend our original example, the fixed probe measuring the temperature on a planetary surface. Consider the same probe attached to a vehicle or rover. The rover allows for simple actions, say: left, right, forward, backward, and stay. If we allow the robot to curiously explore it will, after generating a first model from its observations, begin by choosing actions either randomly or according to some ad-hoc scheme. Based on its observations the reinforcement learning algorithm will modify the likelihood of the actions so as to maximize the reward signal.

In this example the robot will detect a uniform, boring, temperature distribution on the surface, but when it enters a crater it will happen upon a different temperature pattern. This temperature pattern will be novel, and non-random, and therefore interesting. With enough exploration of the phenomenon the robot will make compression progress and enjoy a curiosity reward. This reward signal will then be used by the learning algorithm to increase the likelihood of action sequences that find such regions. As long as the pattern is not fully incorporated into the world model the agent will receive reward for exploration in this area. Continuing, the next time the agent discovers a crater, the temperature profile will again change. A crater of roughly the same size and shape will have the same profile and will generate little reward, exploration will not last long in this area. However, the exploration of a crater of a different shape will remain rewarding. Explorations of craters of different sizes will carry-on until the robot's model is capable of predicting the temperature profile of any crater.

A space probe that is driven by artificial curiosity will lead to the discovery and the modeling of unpredicted phenomena. This probe would not need the direct intervention of human scientists to guide it towards such phenomena, nor would its behavior need to be predetermined and programmed ahead of time. But a probe is typically sent into space with goals other than pure exploration. Often very specific experiments, predefined by scientists on Earth, or constrains on exploring only specific parts of space or certain phenomena are put on the spacecraft. A robot driven solely by exploration is simply not realistic. How then can a curious agent best serve us? Given the goals of a specific space mission we can define a goal to be used in the feedback of our reinforcement-learning algorithm.

Say we want our robots to mine some distant planet for a particular ore. The more ore they mine the better. Our feedback signal is easy to design. But, where is the ore? And what is the best way to mine on this unfamiliar planet? These are questions that the curiosity reward can help answer, by exploring and developing a world model that allows better prediction about the nature of the planet so that the most ore can be mined.

Moreover, how does an agent balance *exploration*, developing a model via an intrinsic reward signal, and *exploitation*, using its model to directly fulfill its goals? This is a non-trivial problem. In general it is hard to balance the exploration and the exploitation of an environment so as to maximize the external reward. It becomes particularly difficult when other constraints, such as limited lifetime and energy usage, are taken into account. Regardless, a reinforcement learning algorithm can learn when to explore and an agent equipped with artificial curiosity gives the agent a guided, open-ended, way to explore so it may better improve its collection of external reward.

Over the years various formalizations of curiosity have been researched by Schmidhuber et al. [20]. Some of them include intrinsic reward based on prediction error [12], world model improvements, differences between prior and posterior beliefs of agents before and after learning new data [22], as well as 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 [14]. There are other approaches as well for adding intrinsic motivation to artificial agents, e.g. [8].

Although it is conceptually clear how a route toward building curious autonomous agents looks like, the current state-of-the-art in machine learning may not yet be sufficient for realizing such behavior in a space probe. We identify two different types of bottlenecks: First, current compressors are mostly limited to specific domains, such as text, images, sound and video. Far more powerful model-building learning methods are required to build compressors that capture the world in the way human scientists do. Second, a reinforcement learner that makes sense of the internal curiosity feedback generated by the compression-based coherence module needs to scale gracefully to larger and more complex environments. This requires progress in the field of reinforcement learning, since current methods are typically limited to small, simple, and low-dimensional tasks.

Current research focuses on finding ways to bring together the measure of compression progress, which is used as an intrinsic reward signal, with general yet practically feasible reinforcement-learning methods. These are yet to be implemented in real world and robotic systems. In our lab we focus on the implementation of these ideas in humanoid robots in real-world environments, as for example, in the E.U. funded project: IM-

CLEVER[1].

5 Curiosity in Science

Compression progress is not only a useful principle for learning machines, such as curious space probes; it also reflects an important aspect of human scientific interest. By improving the subjective compressibility of the history of observations we obtain shorter and simpler descriptions of that history. The regularities that facilitate a short, simple description of the history can be regarded as rules that describe the structure of our observations. Conversely, we can understand our observations in terms of the rules we have thus far discovered. These rules or *compression programs* ultimately form the scientific description of our world. Driven by the desire to find shorter descriptions of their observations, scientists *actively focus* their attention (e.g. build measuring devices, perform experiments) on gathering data that allows them to find or validate better compression programs [15, 18]. Physicists, for example, have traditionally analyzed certain aspects of the world to find simple models to describe their limited observations better than previous models. In essence they are trying to find programs that compress observed data better than the best previously known program. For example, Newton's law of gravity can be formulated as a short program which allows for substantially compressing many observation sequences involving falling apples and other objects. Although its predictive power is limited – for example, while it does not explain the quantum fluctuations of the electrons inside an apple, it still allows for a large reduction in the amount of data required to encode the observations of falling objects, by assigning short codes to events that obey this law. Einstein's general relativity theory yields additional compression progress as it compactly explains many previously unexplained deviations from Newton's laws of motion.

More generally, scientists try to find increasingly compact rules to describe certain aspects of the world that are consistent with rules found eslewhere. for other aspects. However, a description of a system (e.g., a planet's surface) on a certain abstraction level (e.g., particle physics) does not immediately yield insights into all the phenomena related to that system (e.g., craters). Instead, scientists try to achieve further compression and more general explanations by finding rules that allow for shorter descriptions of the concepts known thus far. Such a collection of rules, or a *compression program*, can itself then become a concept to which a certain name is given. For example, planetary scientists obviously do not describe planet surface phenomena on the level of particle physics, but instead use more abstract concepts that allow for shorter descriptions of their observations. They might find it useful to introduce the concept of a crater, based on repeated occurrences of a depression in the planet's surface with a particular shape. The same principle applies for many other concepts we use to describe our world; the concept of a molecule allows for a short description of many aspects of stable configurations of its atoms, the concept of an atom allows for a short description of certain configurations of protons, neutrons and electrons, and so on. Similarly, we can identify more abstract concepts, such as 'turbulence,' for a distribution of motion in a liquid or gas, that is regular

[1] `http://www.IM-CLEVER.eu/`

over a range of scales and allows, once again, for a shorter representation of a significant part of turbulent motion. In this fashion, the amount of compression that can be achieved serves as a criterion for determining the abstraction level on which a phenomenon can best be described. Compression programs for individual concepts on different abstraction levels can again be compared for similarities and grouped by inter-compressibility, potentially yielding further compression. Such an organization of compression programs takes a hierarchical form in which more abstract concepts describe increasingly general relations between concepts on different levels of abstraction. Although compression progress is essential to science, the idea that we should use simple programs to describe our world does not make explicit how to find such programs. As for example shown in [2], simple file compression methods (gzip) can already be used to infer some regularities associated with the concept of 'life,' but not all compression methods used in science might be so straightforward.

The compression of a history of observations not only entails identifying the rules, regularities or models that describe particular physical processes, but also finding the level of abstraction on which physical entities are best represented. Similarly, compression *progress* consists of not only finding shorter rules which describe an ever larger number of observations, but also finding more abstract concepts to which the rules apply. In this fashion, a space probe driven by artificial curiosity could learn to form representations similar to the concepts formed by human scientists. For example, a space probe could learn to represent observations of 'craters' with a short program based on similarities in their shape. After this abstract concept is in place, it could learn to find models for the processes surrounding craters, for example, 'erosion.' Moreover, it could then actively direct its attention or even manipulate its environment, just as scientists perform experiments, to gather observations that allow to further compress its history of observations.

6 Conclusion

We have discussed a formal notion of curiosity within a framework of an agent that, from the interactions with its environment, learns to focus on observations with patterns that were not yet identified. Informal notions related to curiosity, such as complexity, pattern, regularity, novelty, interestingness, were captured in a general computational framework based on compression. Compression programs allow an agent to store its history of observations based on identified regularities underlying those observations. The interestingness of incoming observations can be determined relative to the agent's current ability to compress its history of observations. In particular, interestingness can be measured as *compression progress*. We also introduced a method for measuring the extent to which new data fits old data, called coherence progress.

Many algorithms and methods developed in artificial intelligence and computer science in general, such as gzip, neural networks, pattern recognition or dimensionality reduction, ultimately perform some kind of compression. Of course, most of the existing methods have their own particular limitations: e.g., only successful with specific data types (such as text documents, images or music), do not produce human-readable representations, are overly time or resource intensive. While the current state of the art in machine learning is yet unable to address all these issues in general applications, a major advantage of our approach is that the interestingness of observations is determined *relative* to the pattern discovery ability of the compressor. Patterns that can be easily discovered by a certain compressor soon become boring, while patterns that can never be found by a compressor will also not be interesting. Instead, a curious agent focuses its actions on collecting observations for which its *limited* compressor can find regularities that were not yet discovered in the history of observations thus far. As researchers develop better and more flexible compression methods, the capability of curious artificial intelligence can be extended within the general framework of compression progress.

Artificial curiosity in artificial intelligence is closely related to human curiosity in scientific investigation. Scientists not only try to find regularities in previous observations, they also actively collect new observations that allow them to find even better compression programs. Autonomous exploring probes should resemble human scientists in that regard and use artificial curiosity to discover concepts similar to those found by human scientists.

References

[1] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Occam's razor. *Information Processing Letters*, 24:377–380, 1987.

[2] F. Corsetti and M. Storrie-Lombardi. Lossless compression of stromatolite images: A biogenicity index? *Astrobiology*, 3:649–655, 2003.

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

[4] M. Hutter. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin, 2004.

[5] L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: a survey. *Journal of AI research*, 4:237–285, 1996.

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

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

[8] K. Merrick and M. Maher. Motivated learning from interesting events: adaptive, multitask learning agents for complex environments. *Adaptive Behavior*, 17(1):7, 2009.

[9] J. Poland and M. Hutter. Asymptotics of discrete mdl for online prediction. *IEEE Transactions on Information Theory*, 51(11):3780–3795, 2005.

[10] T. Schaul, L. Pape, T. Glasmachers, V. Graziano, and J. Schmidhuber. Coherence Progress: A Measure of Interestingness Based on Fixed Compressors. In *Fourth Conference on Artificial General Intelligence (AGI)*, 2011.

[11] 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.

[12] 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.

[13] J. Schmidhuber. Artificial curiosity based on discovering novel algorithmic predictability through coevolution. In P. Angeline, Z. Michalewicz, M. Schoenauer, X. Yao, and Z. Zalzala, editors, *Congress on Evolutionary Computation*, pages 1612–1618. IEEE Press, 1999.

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

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] J. Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990-2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230 –247, 2010.

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

[22] J. Storck, S. Hochreiter, and J. Schmidhuber. Reinforcement driven informa-

tion acquisition in non-deterministic environments. In *Proceedings of the International Conference on Artificial Neural Networks, Paris*, volume 2, pages 159–164. EC2 & Cie, 1995.

[23] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.

[24] T. Welch. High speed data compression and decompression apparatus and method, Dec. 10 1985. US Patent 4,558,302.

[25] W. M. Wundt. *Grundzüge der Phvsiologischen Psychologie*. Leipzig: Engelmann, 1874.