On the emergence of science in the frame of realistic observer models

Guido Kruse

Abstract

In this essay I describe how we can use autonomous agents (AAs) from Embodied Cognitive Science (ECS) as models for real observers in order to understand how science may emerge for observers. I model *science* as the ability of observers to use symbolic systems to perform experiments. The methodology consists in designing specific experiments with AAs and application of these experiments to show how we can interpret such experiments as *observers perform experiments using symbolic systems*. According to an idea of Albert Einstein all human-observer notions (in science, daily life etc.) are arbitrarily creations of thinking and cannot be derived inductively from experience [4]. In other words, symbolic systems which are created by observers are only labels for stable sensor input patterns which may originate from environment objects as well as from patterns created by observers themselves. I will introduce a simple experiment with an AA which shows how we can formalize these ideas in an ECS-experiment. For this purpose I introduce various models elements like classifications processes, internal/external categories, correlation of categories and artificial/natural pattern source based categories. I speculate that if such experiments would be performed systematically in ECS with increasing complexity in the long run we will be able to obtain basic propositions for foundational sciences like physics or mathematics from such experiments. This would allow for an experimentally based understanding of how an observer creates a picture of reality.

1 Introduction

I think that if we had a more complete understanding of the structure of observers and their interaction with their environments we would build up science not by starting from empirical facts like c=const. or more formal propositions like Schrödinger equation or invariance requirements. I rather think that in this case basic scientific propositions would be derived from the observer structure and the specific observer-environment interaction process. This approach seems to be natural since all science is "made by observers". And what we know about our specific reflection about our environment (or in other words our *reality*) is shaped by the way we make experiments and "think".

Now the scientific description of the structure of observers and their interaction with their environments is extremely complex and it is hard to find a reasonable degree of abstraction for starting a program like "deriving science from observer properties".

In this essay I will use the Embodied Cognitive Science (ECS) approach to sketch how the mentioned program could already today be studied in ECS-experiments (introductions into ECS are e.g. [1], [2], [3]). Why do I propose to use ECS? I think the Autonomous Agent (AA) models which are studied in ECS together with the quantification of the interaction process of

AAs with its environments have enough structure for modeling key aspects of real observers. These include abilities like autonomous existence, categorization of environment properties, showing reproducible behavior in similar situations, learning and communication with other AAs.

I want to avoid a misunderstanding from the beginning: I suppose that the environment which is composed by objects with certain properties exists *independently* from observers who are embedded in the environment. But there is an observer-model *dependent* interaction with environment. This specific interaction results in what I call *reality* for a given observer-model. In this sense *reality* is an abbreviation for all the information a given observer-model has about its environment.

The distinction between environment as such and reality for an observer becomes clear if we consider that an observer-model is defined by its specific realizations of sensor-/motor-/artificial-neural-systems and the specific observer learning-history. Two observer-models may process completely different *information structures* of the same environment and therefore have different *realities*. Two examples for real observers are humans and ants.

In this essay I will only use concepts which are well defined in an ECS approach and are therefore experimentally accessible. Concepts of physics or mathematics are not used primarily – they should follow from ECS considerations.

What about the essay topic *Connection between Physics and Mathematics*? This topic also has to be understood in context of an observer-model. I will justify this later in greater context.

I start with introducing the concept of an observer-dependent reality and then show how we can understand in principle the emergence of science and therefore of physics/mathematics in this frame.

2 Observer-model based definition of reality

In this section I want to analyze how *observer-model dependent reality* can be understood in context of an ECS model. Since the requirement of science is to make propositions which can be put into context with environment properties the understanding of the *observer-model based definition of reality* should be the starting point for all scientific approaches.

The main point which I want to make in this essay is that it is already feasible with sufficiently complex AA-models to study how science can emerge in the context of an ECS-experiment. The notion *science* is used as an abbreviation for *a model of science* which can be summarized as follows: an observer uses symbolic systems to plan and perform experiments. *Symbolic systems* are modeled as sensor input patterns (e.g.: visual, acoustic input patterns correspond to written/spoken symbols) created by observers themselves and experiments are modeled as connected behavior states of observers.

In the following I will explain more details which are necessary for ECS-experiments which aim at understanding the basic process which underlies emergence of science. The central theme is:

1. Introduction of building blocks of an ECS-experiment
2. Explanation of two important classification processes which can be applied by AAs and which serve as basis for making distinctions in the AA´s environment

3. Based on the classification processes: introduction of external/internal AA categories and its correlations
4. Distinction: environment – reality for an observer
5. Introduction of natural/artificial sensor input (NSI/ASI) patterns and related categories for modeling of environment object properties (NSI) and of symbolic systems (ASI)
6. The correlated system of NSI-/ASI-/ANN-categories is a model for the "knowledge" structure of an observer who is doing science (ANN= artificial neural network)
7. A simple model (experiment) is introduced.

In ECS AA-models and their interaction with its environments are studied (see e.g. [1], [2], [3] for introduction of various AA-models). AA-models from ECS can be used as models for real observers and as tools for studying features of the observer-environment-interaction. From now on the notions AA / observer are used as synonyms for each other. An AA / observer only exist as a specific realization which will be denoted by "AA-/observer-model".

Basic ECS model blocks are:

1. *Environment E* (defined by environment-objects and its properties)
2. *AA-model* (defined by a specific physical setup concerning sensor-/motor-system and artificial neural network (ANN) and the learning history and interaction AA-E history).

The ability to classify/categorize environment properties is a precondition for an AA to be able to fine tune its behavior consistently with its environment structure since the classification process results in the creation of information which can be used by the AA to control its behavior. In ECS different types of classification processes can be considered in dependence on the pattern sources which provide input-patterns for classification.

In the following I will use the notion *pattern* for denoting all kinds of possible inputs/outputs for/from sensor-/motor-/ANN-systems (we can think of high dimensional real number vectors which are processed by target systems). Patterns can be classified by AAs (e.g. by ANN learning). The classification process results in *categories*. The sum of all categories which an AA-model may have created during its life time will be called the *information source* at disposal for that AA-model. In this sense the notion *information* will be used for uncorrelated/correlated categories originating from various patterns sources. Information can be used actively by AAs for controlling behavior. Possible types of categories will be explained below.

I will consider the following two important classification processes:

1. There exists a classification process for AAs called *sensor-motor-coordination (SMC)* which uses the complete physical AA setup and the interaction between E-AA (see [1]). This process differs from standard ANN pattern classification procedures since it does not merely result in static category-node computations. It rather creates *categorical-behavior* (like "circling around an object" – this is the category "circular object" expressed by behavior – there is no simple correspondence to a static "pre-existing" ANN category).
2. AAs can classify patterns by standard ANN-learning which are exclusively produced by its ANNs (patterns not triggered by environment interaction but rather by pattern input from other ANN layers). We may imagine that these patterns are created by

higher neural processes (processes which are not directly coupled to the sensor- and motor-system but rather to other layers of the ANN-system).

I want to briefly sketch the first classification process since it requires an embodied agent model: it is based on an AA-model feature which consists in the AAs ability to emerge *SMC behavior*. SMC enables the AA to structure its sensor input space (e.g. by reduction of the dimension of the sensor input space by performing specific behavior which influences the way how the sensors gather information). Via SMC correlations can be induced between the various sensor channels of an AA-model (e.g.: visual and acoustic channels for human observers). These correlations make learning – adjustment of the connection strengths between the neurons of the ANN - easier and enable the AA to fine tune its behavior for adaptation to its environment (see [1] for explanation of SMC).

SMC can be considered as the basic process for learning environment-categories (classification / categorization is the ability of the AA to make distinctions in its environment by category-specific behavior). A category in this sense is not static information in an ANN but rather a specific kind of AA behavior which requires embodiment: the AA behaves as if it would know a category but in effect the AA-E-interaction produces categorical behavior (which was learned before).

Categorical-behavior caused by stable AA-E interaction patterns (i.e. patterns which are reproducibly emerging if the agent interacts with similar environment objects) is an expression of the AA´s ability to categorize environment object-properties without the need for higher neural process decision making. In ECS for a given AA-model categorical-behavior can be quantified by applying various statistical measures (e.g.: mutual information, entropy, correlation) to the data space created by the sensor- and motor-system. Typical statistical *fingerprints* are found for a given AA-model. These fingerprints reflect the AA´s environment structure as experienced by the concrete AA-model (see [1]).

The second classification process which I mentioned above is the known process from ANN theory: an ANN learns to classify various types of input patterns. After the learning phase is completed certain pattern classes can be recognized. This classification process is different from SMC categorization since it does not necessarily requires embodiment and can be used for classification tasks for higher neural processes.

Based on these two classification processes I will distinguish *external/internal* categories:

1. Categories created from patterns which are direct input for the sensor-/motor-system are *external* categories.
2. Categories created from patterns which are created by higher neural processes are *internal* categories.

External and internal categories can be correlated. This only requires sufficiently complex ANNs and physical setups. I will use the term *correlation of categories* in the following sense using *time* for denoting the time steps which refer to the update process of sensor-/motor-/ANN-system-states of an AA-model:

1. *Correlation of internal-internal categories:*
   Categories are *correlated* if the related category nodes of an ANN are reproducibly activated according to a temporal order relation (distance in time with order: before/after/simultaneously).

2. *Correlation of external-external categories:*
   If we define categorical behavior as a temporal related closed sequence of characteristic environment induced sensor-/motor-states than two distinct categorical behavior patterns are *correlated* if they are reproducibly connected via a defined temporal order relation.

3. *Correlation of external-internal categories:*
   Mixed categories are *correlated* if category nodes related to internal categories do reproducibly initiate categorical behavior states (or vice versa) via a defined temporal order relation.

In general only *AA-model-specific* categories can be learned and used for behavior control since different AA-models have different physical setups including different ANN structures. This can be observed by the different behavior states which emerge from different AA-models in identical environments. It follows that the information about the AA´s environment which is the AA´s *reality* depends on the AA-model. Even two identical AAs from the same AA-model may have different *realities* in an identical environment because their learning histories may differ.

Based on this fact we again stress the following distinction because of its epistemological relevance:

1. The *Environment* exists independently from an AA-model and is composed by objects which have certain properties. The interaction E-AA may influence both E and AA properties.
2. The notion *Reality* is used as an abbreviation for the *information* i.e. the sum of correlated/uncorrelated external/internal categories which an AA-model has created during its life time from patterns originating from the coupled system *S = E + AA-model + interaction [E-AA]*. *Reality* in this sense is not static and may develop during an AA´s life time.

Since information about an environment necessarily includes information about the *AA-model* itself and the specific *E-AA-interaction* only a mixture of information originating from *E + AA + interaction [E-AA]* can be in principle described by science. This fact justifies the *observer-model based definition of reality* which can be scientifically modeled in ECS experiments in contrast to an *objective definition of reality* which cannot be given a scientific meaning.

Here are some examples for pattern sources originating from:

1. *environment (objects):*
   - object reflects light which is received by a visual sensor channel like a AA´s camera
   - E-temperature which can be quantified by a temperature sensor
2. *interaction of E-AA:*
   - a series of proximity sensors which measure the distance to an object enable e.g. circling behavior around a circular environment object (see [1]).
   - patterns are produced while an agent is interacting with an object from the environment and are collected via sensor-/motor-system with active SMC of the AA-model. ANN learning/adjustment makes the behavior reproducible.
3. *AA-model only:*

- No direct external sensor input originating from 1./2. is required. These patterns are created by sensor-/motor-/ANN-systems as such without AA independent environment E involved.
- Examples are e.g. speech for acoustical sensors, symbol patterns for visual sensors and category-patterns only from ANN which again serve as input pattern for ANN-structures.
- In principle we can add all kinds of higher neural processes to the ANN. These processes can interact with each other and with processes directly coupled to the sensor-/motor-systems. The classification of output patterns from higher neural processes as the basis for information from internal categories was mentioned above.

All patterns originating from (1. – 3.) can be classified by an AA-model using the two classification processes described above. The arising categories can be correlated and act as sources for *information* which can be used by the AA to control its behavior.

Real observers use symbolic systems to plan and perform experiments which can be identified with connected complex behavior states of an AA. The question is: how can we model *AAs using symbolic systems* in the frame of ECS? It turns out that this is already possible using the classification processes explained above together with sufficiently complex AA-models. The main idea is that *symbolic systems* can be identified with patterns created by the AA itself (I will call these patterns *artificial sensor input patterns* – ASI-patterns). Examples are:

- the AA creates visual patterns as input for its visual sensor channel (formulas are an example in the case of human observers)
- the AA creates acoustic patterns as input for its acoustic sensor channel (speech is an example in the case of human observers)
- the AA correlates visual and acoustic patterns for both visual and acoustic sensor channels (speaking about formulas is an example in the case of human observers).

I proceed with analyzing how a given AA-model could classify its environment by using correlations between categories created from “artificial” sensor input (ASI) and “natural” sensor input (NSI) patterns (since this models a realistic observer with the abilities for communication and information enrichment of environments by creating ASI). I use the following definitions:

1. All patterns which are created by an AA-model´s motor-system for direct input for the AA´s sensor system are called *ASI-patterns*. Categories from ASI-patterns – *ASI-categories* – are *external* categories.
2. All patterns which are originated by environment objects are called *NSI-patterns*. Categories from NSI-patterns – *NSI-categories* – are *external* categories.
3. All patterns which are created by an AA-model´s ANN-system called *ANN-patterns*. Categories from ANN-patterns – *ANN-categories* – are *internal* categories.

Examples:

For 1.: Using its motor-system AAs can create patterns for sensor input which would not exist without the AA as part of the environment (AAs enrich its environment with “artificial” information). These signal-patterns include e.g. written symbols (ASI for visual sensor channel) or spoken symbols (ASI for acoustic sensor channel). Examples for these patterns are visual and acoustic pattern created by an AA itself (handwriting, speech).

For 2.: All patterns which exist independently from AAs. These signal-patterns originate from environment objects and can be input for various AA sensor channels. Reflected light for the visual sensor channel is an example.

For 3.: Imagine an ANN-structure which consists of separable and connected ANNs. Each single ANN may produce outputs which can be classified by other parts of the ANN-structure.

AAs can classify ASI patterns in the same way as NSI patterns since both pattern sources provide direct input for the AA´s sensor system. The resulting categories resp. information can be correlated among each other and with ANN-categories and processed by the AA. With the help of this mixture of correlations an AA can coordinate / fine tune its behavior autonomously or together with other AAs (communication) consistently with the environment structure. This kind of categorical-behavior which can be partially controlled by ASI-categories (symbolic systems) is the basis for planning/performance of experiments by an AA in its environment.

Now we already have enough model elements for performing ECS experiments, which are targeted on understanding how *science* may emerge (let's call this type of experiments "quantitative epistemology" (QE) - experiments). A minimal experimental setup requires an environment with objects and an AA. Then the experiment consists in recording of data spaces related to sensor-/motor-/ANN-states. These data spaces can be analyzed using e.g. statistical or information theoretic measures. Statistical fingerprints from different data spaces can be identified and correlated (see [1]).

A QE-experiment is characterized by the fact that AAs itself *create* input patterns (ASI) for its sensor-system in order to *label* reproducible behavior states and NSI-categories which emerge from E-AA interaction in similar environments. QE-experiments are models for real observers who are using symbol systems to plan and perform a real experiment (like in physics).

ASI-categories in this connection can be notions from experimental physics or mathematics – all notions correspond in our model to ASI-categories. Direct experimental results like reading off a scale correspond to NSI-categories. Resulting ASI-/NSI-categories can be correlated. Now we may ask how it is possible to correlate ASI-/NSI-categories. It may result from QE-experiments that output from higher neural processes is required to correlate ASI-/NSI-categories via ANN-categories. This could be a toy model for "thinking".

Using the experimental strategy sketched so far we could systematically investigate how observers build up "knowledge" about their environments. The knowledge-structure is much more complex than it is usually assumed in physics papers since it is build up from AA-model dependent correlations of ASI-/NSI-/ANN-categories which in turn result from various classification processes applied to various input pattern sources.

In the following table we sketch a toy model setup for a QE-experiment: an AA performs an experiment using an ASI-category. The experiment consists in distinguishing two objects (rectangle cross section *R*, circle cross section *C*) and pushing *C* over a certain distance after having detected a specific ASI-category which serves as a label for *C*. we can interpret this experiment like: an observer uses a symbolic system to perform an experiment.

AA behavior states	Classification Process	ASI-pattern	NSI-pattern	ANN-pattern	Correlation	Interpretation
1. learn categorical-behavior *circling around object C* (see [1])	SMC resulting in specific environment-fingerprints (see [1])		x	x	NSI/ ANN	
2. learn recognizing *C* with the help of the specific fingerprint from *circling around object C* behavior – this results in category node for *C*	ANN-learning based classification		x	x	NSI/ ANN	Object identification
3. learn creating arbitrarily acoustic signal (e.g.: *click/click*) as label for *C* after having recognized *C*	SMC resulting in specific environment-fingerprints	x		x	ASI/ NSI/ ANN	Symbolic system for labeling *C*
4. Learn pushing *C* for 10cm after having heard *click/click*	SMC resulting in specific environment-fingerprints	x	x	x	ASI/ NSI/ ANN	the experiment

The table displays the succession of behavior states which result in performing the experiment by the AA. For each behavior state I show the classification process which is involved together with the sources of input patterns. The last column shows the interpretation in context of a QE-experiment. The AA learns behavior according to steps 1.-4. from the table. The steps are based on each other (e.g.: step 4. requires steps 1.-3.). Then the AA is placed in an environment equipped with objects *R/C*. In case of recognizing *C* the AA performs the experiment after hearing *click/click*: the AA uses a "symbolic system" to control an experiment. If the AA recognizes *R* then no experiment will be performed because in this case the AA learned in step 4. to create another "symbol" – like *click/click/click* - for labeling *R*. Only the "symbol" *click/click* triggers the experiment. In the experiment above the AA uses a very simple symbolic system - *click/click* - to trigger the experiment. More complicated symbolic systems (based on ASI) together with communication with other AAs to trigger more complicated experiments would make QE-experiments more realistic. But in principle such experiments are feasible today.

I want to finish this section by stating that there is a substantial deviation between the concept of the knowledge-structure of an observer like explained in this essay and the one which is sometimes used in physics to speculate about foundations of quantum physics (e.g.: "observer-model" consists in the entity which reduces a state superposition). Apart from the oversimplifications made sometimes in physics concerning properties of observers I think that the basic methodology is not appropriate: we should not model observers mathematically as part of theories of physics/science since we neglect in this way many relevant observer-model properties. The correct methodology should be to *derive* foundations of science *from* a realistic observer-model taking environment interaction into account.

3 Speculation

The *connection between physics and mathematics* is a system of correlated NSI-/ASI-/ANN-categories in the context of the human observer-model. This system allows for complex behavior coordination which results in planning/performing experiments. The correlated behavior states called *experiment* allow an observer to reproducibly create sensor-input states which can be categorized and correlated with new/existing NSI-/ASI-/ANN-categories. In this way an AA increases its "knowledge" about environment and about itself and enlarges its scope of *reality*.

It is important to understand that according to the simple observer-model and the toy model for the QE-experiment from section 2 it is already possible to describe *quantitatively* in experiments the complex behavior as well as higher brain processes both resulting in some kind of specific behavior called "doing science". An important aspect of such experiments would be the systematic investigation of the interplay between categories resulting from artificial sensor-input-patterns (ASI) which were created by the AA (e.g. acoustic/visual patterns) and natural sensor-input-patterns (NSI) resulting from environment objects. This kind of analysis would be a starting point for a scientific description of a human observer`s knowledge gaining process.

It is interesting to try to yield some basic propositions with empirical evidence for fundamental sciences (like physics) in the light of ECS QE-experiments. What is the most basic criterion for a proposition being fundamental for physics: it has to be consistent with the very definition of *reality* (see our description in section 2). From a technical point of view we have to find a mathematical formulation for a *reality consistency condition* for physical laws (we can think of an invariance condition etc.). So called "elements of reality" would then be assigned to certain ASI-/NSI-/ANN-category-correlations according to the model from section 2.

From a quantitative epistemological point of view quantum mechanics as well as classical mechanics and more generally all branches of physics are expressions of an observer-model´s "knowledge" (which is the set of correlated categories described above). There is in principle no "1:1" correspondence between ASI-categories and something objectively real (objects from environment) like it is suggested e.g. in classical mechanics. Closest to environment properties are NSI-categories according to our observer-model. Therefore, since science is made by observers, we have to investigate the "knowledge"-structure of an observer-model. This is possible using realizations of an observer-model in ECS. Based on this analysis basis propositions for physics should be derived.

4 Summary, Conclusion and Outlook

I sketched how it is possible to design experiments aiming at understanding of how science may emerge for human observers using embodied cognitive science approach. It is already possible today with existing technology to perform such experiments and to search systematically for processes underlying the human knowledge gaining mechanism. Once we have understood this mechanism it should be possible to clarify the epistemological foundations of science experimentally. This methodology opens a wide field of research which will result in interesting fields of applications and insights. E.g.: the systematic investigation of autonomous agents which are able to create correlated *symbolic systems* will result in setup of new AA-models with optimized abilities for behavior control in specific environments.

To conclude: if this approach turns out to be meaningful than its systematic application would lead to a deepened understanding of foundations of science and of how human culture in general emerges.

References

[1] Rolf Pfeifer and Christian Scheier, *Understanding Intelligence,* (The MIT Press, Cambridge Massachusetts, London, England, 1999)

[2] Rolf Pfeifer and Josh Bongard, *how the body shapes the way we think : a new view of intelligence,* (The MIT Press, Cambridge Massachusetts, London, England, 2007)

[3] Fumiya Iida, Rolf Pfeifer, Luc Steels, Yasuo Kuniyoshi (Eds.), *Embodied Artificial Intelligence*, (Springer-Verlag Berlin Heidelberg 2004)

[4] Einstein, Albert: Mein Weltbild, erstmals 1934 von Rudolf Kayser herausgegeben; bearbeitete/ergänzte Neuauflage hrsg. von Carl Seelig, Europa Verlag, Zürich/Stuttgart/Wien 1953 - auch: Frankfurt/M., Berlin 1965, ebenfalls als Ullstein Taschenbuch ISBN 3-548-36728-3, 2005