

# Intelligent description and the principle of free energy

*Friedrich Sösemann 5/2024*

"You have to make things as simple as possible.  
But not easier."

Albert Einstein

## Terms

*Quality, quantity, state, variety, frequency, dependency, causality, correlation, relation, assignment, information, object, subject, description, truth, knowledge, intelligence, principle of free energy.*

## Abstract

Information measures the dependency between states, knowledge that between object and subject states and intelligence that between subject states. Descriptions store object states.

Friston's free energy principle is intelligent, combining physics, computer science and biology, but is not new.

## Problem

Descriptions are simple, but not too simple if they are true and intelligent.

In the following, descriptions themselves are described as simply as possible, but still precisely. They and their measures of information, knowledge and intelligence are derived from a given worldview.

Afterwards, an attempt is made to classify and evaluate Karl Friston's principle of free energy.

## Quality and Quantity

The *world* consists of different *quantities* in the *space*. The quantities change over *time* within the scope of their *quality*.

Quality and quantity are mathematically described by sets whose elements exclude each other.

Examples include *properties* and their values, *alphabets* and their letters or *number systems* and their numbers /1/.

*Continuous* Qualities have an infinite number of ordered quantities. *Discrete* ones usually a finite number, which can also be unordered (*nominal scales*).

Natural numbers are discrete, the real ones are continuous. The property color, with the values red, green or blue, is discrete and nominal, as a *spectrum* it is continuous and *ordinal*.

*Entities* are places or areas from several neighboring places. They are usually associated with a few qualities.

The totality of the quantities of all existing or considered /2/ qualities are called the *state* and is mathematically shown as *tuple*.

Examples of this are numbers, *words* or the representation of states as points in *state space* and in *tables*.

## Frequency and Variety

The size of the chosen area and its qualities determine the number of entities. These and the number /3/ of their quantities determine the range of possible states.

The number of possible different states  $N_M$  is the product of the numbers of the quantities  $N_m$  of all  $M$  qualities considered:

$$N_M = \prod_m^M N_m.$$

The number of states actually present, i.e. the number of entities  $H_M$ , is calculated from the number of possible states and their frequency  $h_{nm}$  /4/:

$$H_M = \sum_m^M \sum_n^{N_m} h_{nm}.$$

The absolute frequency  $h_{nm}$ , divided by the number of entities  $h_m$  of quality  $m$  gives their *relative frequency* or *statistical probability*  $p_{nm}$  /5/:

$$p_{nm} = h_{nm} / h_m.$$

The number of quantities also represents the *variety* of a quality. A measure /6/ of variety  $H$  should be additive for several qualities, but should take into account that the number grows multiplicatively and should result in the value zero for only one possible state. This is fulfilled by the *logarithm* of the number of states. Its base two then gives bit as the unit of measurement:

$$H_M = \sum_m^M H_m = \text{ld} \prod_m^M h_m.$$

A quality with equally frequent occurrence of its quantities is perceived as more diverse than if one occurs predominantly and the other only rarely /7/.

In order to reflect this in the measure  $H_m$ , the relative frequency  $p_{nm}$  of occurrence of the entities is taken into account and their sum is weighted accordingly /8/:

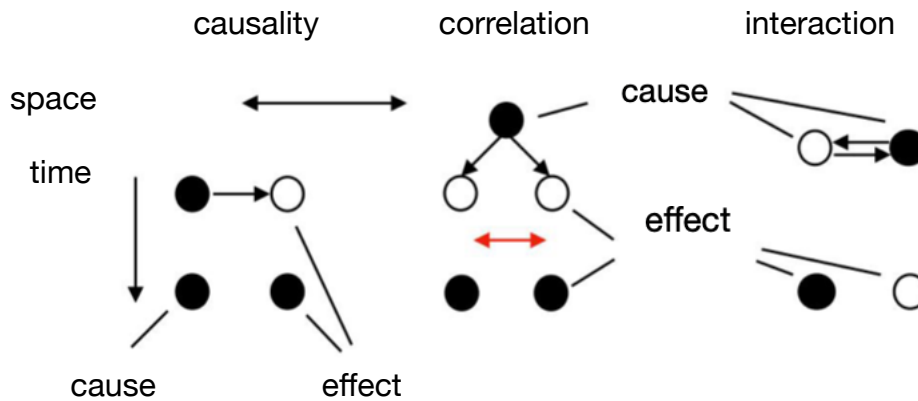
$$H_m = \sum_n^{N_m} p_{nm} \text{ld } 1/p_{nm}.$$

With equal distribution, the diversity becomes maximum and goes into  $H_m = \text{ld } N_m / 9/$ .

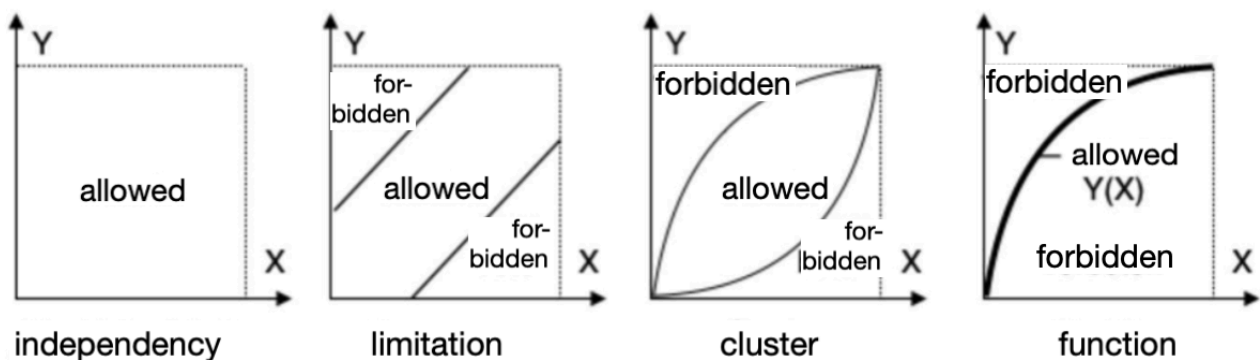
## Dependency and Information

Qualities interact with each other: The quantities of their quality exclude each other and the quantities of different qualities depend on each other, favoring or inhibiting each other. In this way they create *correlation*, *causality* or *interaction*.

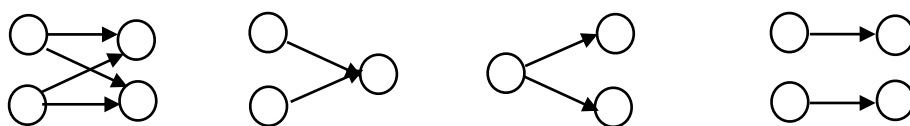
The causality is *asymmetrical*, distinguishes cause from effect. The correlation is *symmetrical* and is caused by third party causality. The symmetrical interaction is mutual causality.



The dependence /10/ can be defined as attraction or repulsion /11/, as permitted or forbidden *assignments* of quantities /12/ and in the form of natural *laws* or more artificial regulate and *definitions*:



Visually, the dependence can be represented by graphs:



Assignment: completely      unique      ambiguous      one-to-one  
 Dependence: independent      fuzzy      randomly      determined

mathematically through *relations*:

$$X = \{ x_1, \dots, x_{N_x} \}, Y = \{ y_1, \dots, y_{N_y} \}, R(X,Y) \subseteq X \otimes Y,$$

these are subsets of the cross product of sets, or practically by *tables* /13/.

The *information* /14/ is the measure of the dependence of states, the difference between the dependent and the independent, maximum diversity /15/:

$$I = H_{\max} - H = \text{Id } N_{\max} / N.$$

This is the case when selecting one of N exclusive quantities:

$$I = \text{Id } N,$$

with M interdependent qualities over the numbers:

$$I = \text{Id } N_1 * N_2 * \dots * N_M / N_{12..M}$$

or more detailed about the frequencies:

$$I = \sum_{n_1}^{N_1} \sum_{n_2}^{N_2} \dots \sum_{n_M}^{N_M} p_{n_1 n_2 \dots n_M} * \text{Id} ( p_{n_1 n_2 \dots n_M} / p_{n_1} * p_{n_2} * \dots * p_{n_M} ).$$

## Subject and Description

*Objects* are entities whose state is relatively constant and differs from that of their surroundings. *Subjects* are objects that can store past states /16/.

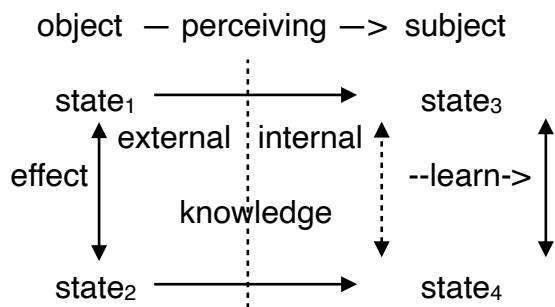
Objects are passive, react, are treated and perceived. Subjects are active, act, behave and perceive /17/.

*Descriptions* are passive, like objects, store, like subjects, and represent external states in internal ones.

All three contain information, so they must have at least two different quantities /18/. Subjects and descriptions realize *intentionality* (Junghans 1999), pointing to other entities.

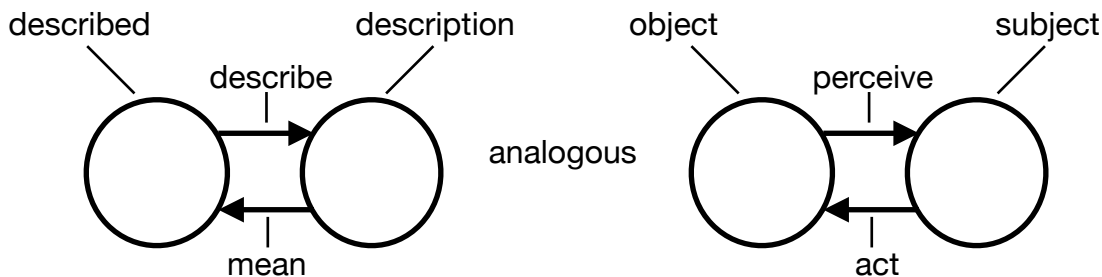
Objects are perceived by subjects by assigning external states to internal ones /19/. The dependence of the external states thus transitively limits the internal ones.

Learning then stores the perceived combinations, supplementing indirect with direct associations, replacing correlation with causality /20/:



The learned redundant assignments increase the efficiency of subjects; assignments that have been learned once do not have to be determined multiple times from the environment.

Descriptions are objects that represent the objects described. Their states passively map and store external states, transforming time into space.

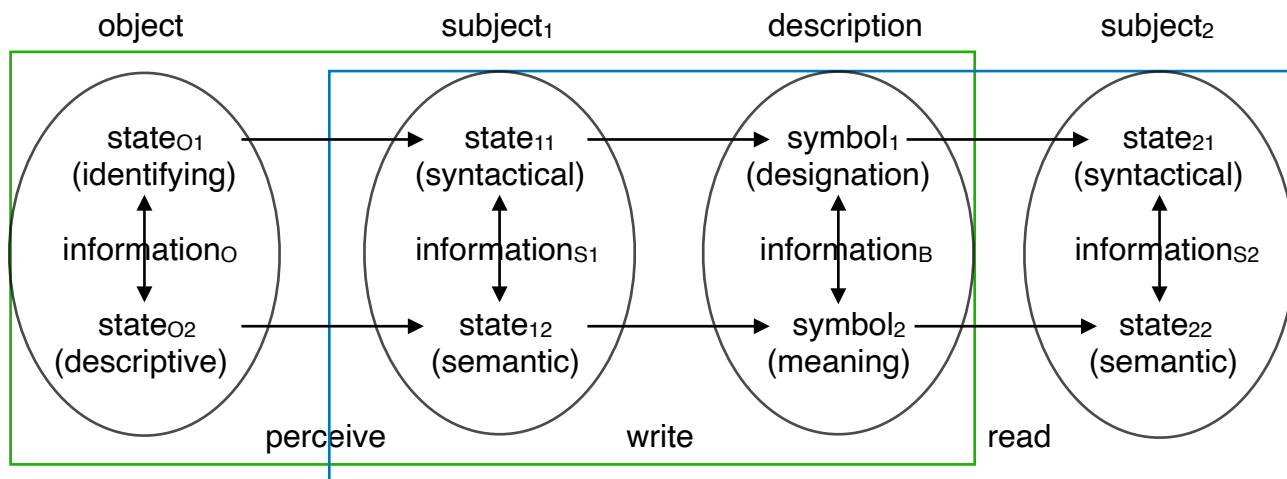


Descriptions are usually sequences of symbols that are assigned to each other via their names.

Listening and reading, speaking and writing are sequential, but what is described is usually a network of entities and relationships. Therefore description follows a path through the network, there are *recursive definitions*, *top down* or *bottom-up* approaches / 21/.

Errors, contradictions, completeness or probability are not characteristics of reality, of what is described, but of the description alone.

Descriptions are used to store knowledge and share it between subjects:



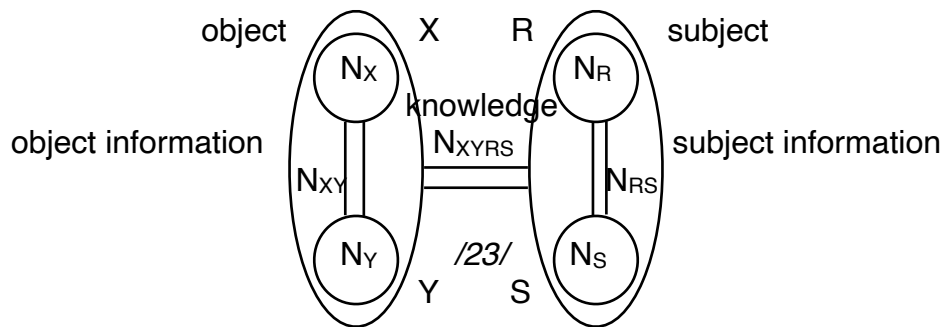
Descriptions are states of a *language*. Their *syntax*, *semantics* and *pragmatics* are intersubjectively agreed definitions /22/. The syntax embodies the structure, its information, the semantics, the meaning, its knowledge, and the pragmatics, the efficiency, its intelligence:

level	term	measure	description
0	value, impact	$N_1, N_2, N_{12}$	words
1	information	$I_{XY} = \text{ld}(N_X * N_Y / N_{YX})$	syntax
2	knowledge	$I_{XYRS} = \text{ld}(N_{XY} * N_{RS} / N_{XYRS})$	semantics
3	intelligence	$I_M = \text{ld}(N_1 * N_2 * .. * N_M / N_{12..M})$	pragmatics

## Truth and Knowledge

The dependency between two areas with dependent states, such as object and subject, can be interpreted as *knowledge*; it is information about information:

$$I_{OS} = \text{ld}(N_{XY} * N_{RS} / N_{XYRS}).$$



The knowledge between the two areas plus the internal information of the areas results in the information about the entire area:

$$\begin{aligned} I_{XYRS} &= I_{XY} + I_{RS} + I_{(XY)(RS)} \\ &= \text{ld}(N_X * N_Y / N_{XY}) * (N_R * N_S / N_{RS}) * (N_{XY} * N_{RS} / N_{XYRS}) \\ &= \text{ld}(N_X * N_Y * N_R * N_S / N_{XYRS}). \end{aligned}$$

Information as a measure of dependency should not be confused with the content of information, the colloquial “information” /24/. So knowledge is necessary, but only *true* knowledge is sufficient.

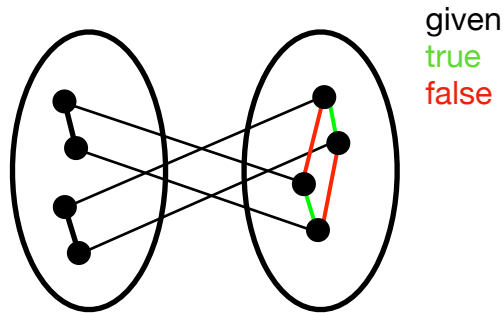
Subjects possess true knowledge as an internal model of objects or their environment. Models of subjects  $S(R)$  are true with respect to the object  $Y(X)$  if for their functions

$$Y(S(R(X))) = Y(X)$$

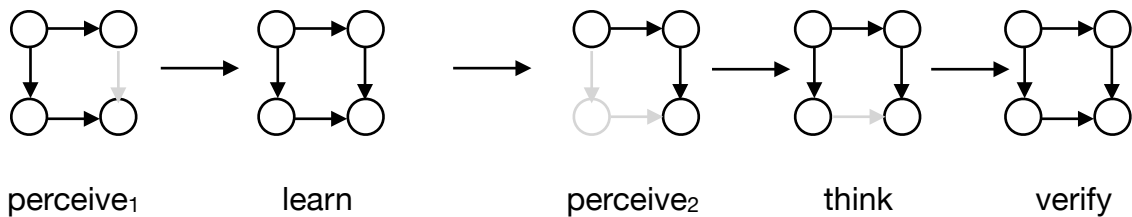
applies, where  $R(X)$  and  $S(Y)$  are the perceptions,  $X(R)$  and  $Y(S)$  are the actions of subjects /25/.

Truth requires structure-preserving assignments, *isomorphism*:

$$I_{\text{true}} = I_{\text{d}} \ 4/2 = I_{\text{false}}$$



The truth of knowledge is *verified* when the predicted quantity actually present:



Knowledge does not require consciousness. With the thermostat, the external information, heating increases the temperature, is firmly linked to the internal information, if the temperature is too low, then heat more.

There are four players in descriptions: the object or environment O, the writing subject S<sub>1</sub>, the description B and the reading subject S<sub>2</sub>. This leads to three connected knowledge relationships, I<sub>OS1</sub>, I<sub>S1B</sub> and I<sub>BS2</sub>, which *transitively* result in the desired knowledge I<sub>OS2</sub>:

$$I_{OS2} = I_{OS1} + I_{S1B} + I_{BS2} = I_{\text{d}} \ N_O * N_{S1} * N_B * N_{S2} / N_{OS1BS2}.$$

If all three parts of knowledge are true, then so is the transitive overall knowledge. In fictions, the object is missing, and in the case of errors, misunderstandings or lies, at least one assignment is wrong.

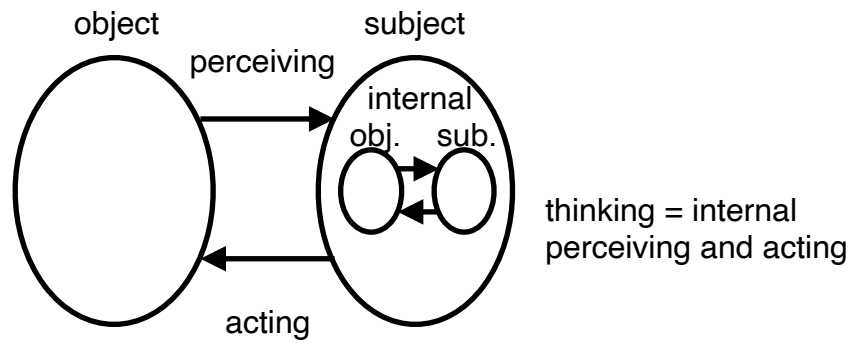
Contradictory knowledge ( X=x<sub>i</sub> & X=x<sub>j</sub> with i≠j ) in descriptions does not represent external information (*disjoint* quantities). Contradictions between objects, descriptions and subjects can be corrected, *dialectical* Contradictions in objects or the environment lead to cycles /26/.

Precise descriptions clearly assign states, thus avoiding false generalizations, misunderstandings or questions.

Descriptions are *understandable*, if description states can be mapped to reader states.

## Effort and Intelligence

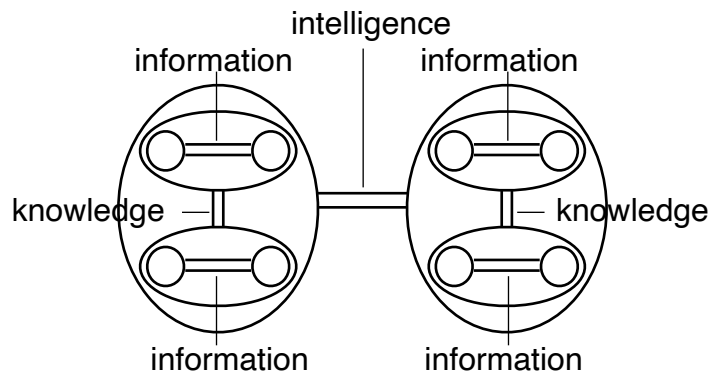
Learning creates knowledge, thinking uses knowledge, replacing external with internal navigation. As a result, the subject acts more efficiently and safely than directly in the environment /27/.



Intelligence is the measure of efficiency of thinking by reducing effort. This is achieved by learning the dependencies between the knowledge elements. The internal structure is improved by transforming the knowledge set into a knowledge network.

*Intelligence* is the measure of dependence on knowledge, the reduction of the number of connection options for all knowledge elements:

$$I_{XYRSM} = \text{Id} ( N_{XYRS1} * N_{XYRS2} * .. * N_{XYRSM} / N_{XYRS12..M} ) .$$



Intelligence is therefore another level /28/ of dependencies and requires an additional spatial dimension:

information	1 dimension		2 qualities
knowledge	2 dimensions		meta-information 4 qualities
intelligence	3 dimensions		meta-knowledge 8 qualities

Example of the three dependency measures:

information	$l_{o1} = \{ (x_1, y_1), (\bar{x}_1, y_2), (\bar{x}_2, y_1), (x_2, y_2) \}$	$= \text{Id } 4/2 = 1 \text{ bit,}$
	$l_{o2} = \{ (\bar{x}_1, y_1), (x_1, y_2), (x_2, y_1), (\bar{x}_2, y_2) \}$	$.. = 1 \text{ bit,}$
	$l_{s1} = \{ (r_1, s_1), (\bar{r}_1, s_2), (\bar{r}_2, s_1), (r_2, s_2) \}$	$.. = 1 \text{ bit,}$
	$l_{s2} = \{ (\bar{r}_1, s_1), (r_1, s_2), (r_2, s_1), (\bar{r}_2, s_2) \}$	$.. = 1 \text{ bit,}$



$$\begin{array}{ll}
\text{knowledge } l_{OS1} = \{ (l_{O1}, l_{S1}), (\cancel{l_{O1}}, \cancel{l_{S2}}), (\cancel{l_{O2}}, \cancel{l_{S1}}), (l_{O2}, l_{S2}) \} & \dots = 1 \text{ bit,} \\
l_{OS2} = \{ (\cancel{l_{O1}}, \cancel{l_{S1}}), (l_{O1}, l_{S2}), (l_{O2}, l_{S1}), (\cancel{l_{O2}}, \cancel{l_{S2}}) \} & \dots = 1 \text{ bit,} \\
\text{intelligence } l_{OS12} = \{ (l_{OS1}, l_{OS1}), (\cancel{l_{OS1}}, \cancel{l_{OS2}}), (\cancel{l_{OS2}}, \cancel{l_{OS1}}), (l_{OS2}, l_{OS2}) \} & \dots = 1 \text{ bit.}
\end{array}$$

*Storage* past states distinguishes subjects from objects. Thinking, as the use and processing of knowledge, presupposes that the internal states that embody the knowledge are preserved over time.

Memories transform time into space, are entities with independent constant states /29/.

*Thinking*, like *deduction*, *induction* or *compression*, is knowledge processing. Deduction draws conclusions from implicit to explicit knowledge. Induction creates new knowledge from existing knowledge. Induction increases, compression /30/ reduces the number of knowledge states in the subject.

The effort, *work* and costs for storing and processing information should be as low as possible. It depends on the amount and type, as well as on the *relevance* and *redundancy* /31/ of the stored knowledge.

Descriptions also store knowledge. Intelligent descriptions are relevant, low in redundancy and explicit:

Relevant descriptions affect and improve the reader's condition and are often the result of a search and selection of the description itself.

Reading the same description multiple times or multiple identical statements in a description result in redundancy. It can be avoided by simply appearing and specifying the number of repetitions. But redundancy can avoid errors in transmission or consolidate learning, "Repetition is the mother of study".

*Explicit* Descriptions save the reader to deduce knowledge from *implicit* ones.

Is the principle of minimum free energy an intelligent description?

## Principle of free energy

The *principle of free energy* by Karl Friston wants to define life, mind and AI and is currently being discussed, advocated /32/ and criticized /33/.

The basis is organisms that have to maintain their vital conditions in a changing environment.

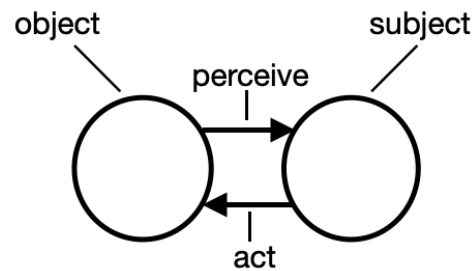
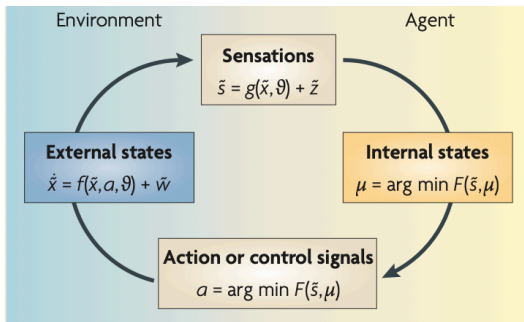
For that brains constantly make predictions about the world /34/ and draw *Bayesian conclusions*, to avoid surprises as much as possible. Because energy is saved by paying *attention* is only activated when something new or unusual happens.

Friston now applies the existing *minimum principle of free energy* to the interface between organisms and their environment.

The principle states that because the total energy  $E$  remains constant, the bound  $E_{\text{bound}}$ , according to the second law of heat theory, tends to a maximum, and the free energy  $E_{\text{free}}$  must tend to a minimum:

$$\begin{array}{rcc} \text{energy} & = & \text{heat} + \text{work} \\ \text{constant} & & \text{maximum} \quad \text{minimum} \\ E & = & E_{\text{bound}} + E_{\text{free}}. \end{array}$$

In doing so, he structures the interaction between the subject and his environment (see picture left from (Friston 2010) ) as above is done (see right):



He considers Bayesian reasoning, learning by prediction and minimal free Energy as belonging together, however:

1. Bayesian learning requires a statistical perspective, but other forms are possible (see above).
2. Learning through prediction already requires an internal environmental model. However, this can be achieved step by step through predictions with possible corrections. But the learning effort will only be reduced if re-learning is more efficient than new-learning.
3. The term *free energy*  $F$  was coined by Helmholtz in 1882. With  $U$  as the internal energy,  $T$  as the temperature and  $S$  as the entropy applies:

$$F = U - T \cdot S.$$

The minimum principle of free energy has also been known for some time (Callen 1985). The connection between information and physical entropy was already described by Leon Brillouin in 1956 (Brillouin 1956) and discussed by the author in 1975 (Sösemann 1975). Karl Friston now uses his “principle of minimal free energy” (Friston 2010) to define life, mind and artificial intelligence:

$$\begin{array}{rcc} \text{energy} & = & \text{heat} + \text{work} \\ E & = & E_{\text{bound}} + E_{\text{free}}. \end{array}$$

But he also points out other interpretations /35/ and links it to Bayesian conclusion, information theory and cognition:

Statistical physics	Bayesian inference and information theory	Cognitive interpretation
Minimize variational free energy	Maximize model evidence (or marginal likelihood); minimize surprisal (or self-information)	Perception and action
Minimize expected free energy; Hamiltonian principle of least Action	Infer the most likely (or less surprising) course of action	Planning as inference
Attain nonequilibrium steady-state	Perform approximate Bayesian inference	Self-evidencing
Gradient flows on energy functions; gradient descent on free energy	Gradient ascent on model evidence; gradient descent on surprisal	Neuronal dynamics

Better than the explanation about energy seems to be the explanation about free and bound diversity, the information:

$$\text{diversity} = \text{bound diversity} + \text{free diversity}$$

$$\text{Id } N_X * N_Y = \text{Id } N_X * N_Y / N_{XY} + \text{Id } N_{XY}.$$

The pattern is the same, but the terms are more appropriate. Minimizing free energy is eliminating the errors of the internal model.

A subject with optimal and true knowledge of the world no longer needs to learn or change the environment; it is in equilibrium with the environment. No further effects are then required across the object/subject boundary. The minimum free energy corresponds to the maximum knowledge.

Subjects differ from objects or their environment through greater complexity, the ability to store data in order to represent the world even in the past, and they correlate more strongly than just interacting objects, in short: they have true knowledge.

An *isomorphism* between areas of knowledge indicates a higher-level structure or *hierarchy* layers that build on each other.

*Analogy* should be fruitful; structures and laws of the more developed area should be transferred to the other.

With hierarchy, biology would have to be derived from physics, taking *emergence* into account in higher layers.

Friston's principle is still far from both, but can be seen as an incentive to break down rigid views.

## Notes

The notes have been moved out of the main text to make reading it more fluid. *Italicized* links make it easier to jump back and forth. Other italicized text passages refer to external sources, such as Wikipedia entries or online articles.

**1:**

Space and time are not qualities, but containers or arrangements of quantities. In *Newton*'s absolute conception of space, it is a container. In *Leibniz's* relational conception, quantities are what make up space. Thus, places are derived, redundant quantities; there are no places without quantity.

**2:**

Whether the qualities and quantities, and thus the entities, *objectively* exist or subjective, the philosophical question is whether *naturalism* or *constructivism*, is valid. It is not intended to answer that here.

**3:**

Objects/events in a space/time area, the elements of a set, the frequency of the elements of multisets or the properties of a object in product sets can be counted:

$$\begin{aligned}
 | ( 0, 1, 1, 0, 0, 0, 1, 1 ) | &= 8 \quad \text{entities in total,} \\
 | \{ 0, 1, 1, 0, 0, 0, 1, 1 \} | = | \{ 0, 1 \} | &= 2 \quad \text{values in set,} \\
 | ( 0, 1, 1, 0, 0, 0, 1, 1 ) | = | \{ (0, 4), (1, 4) \} | &= 2 \quad \text{frequencies in multiset,} \\
 | \{ (0, 1), (1, 0), (0, 0), (1, 1) \} | = | \{ 0, 1 \} \times \{ 1, 0 \} | &= 4 \quad \text{states in product set.}
 \end{aligned}$$

Counting continuous quantities is problematic. Where are the boundaries drawn between the elements to be counted? Intersubjective agreements, such as the definition of the frequency ranges of colors, or a given measurement accuracy are included in the number.

**4:**

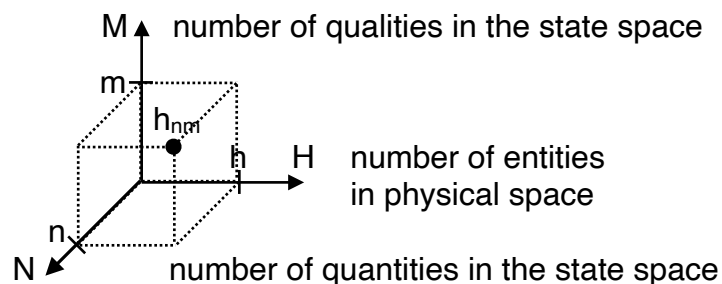
For frequencies  $h_{nm} > 1$  to occur, the number of entities must be greater than the number of different states:  $H_M > N_M$ .

In order to find all occurring quantities of the qualities under consideration, a large space, the "world", must be searched over a long time. In areas that are too small, possible states do not occur:  $h_{nm}=0$ .

If the number of qualities  $M$  taken into account is increased, the number of possible states  $h_{NM}$  increases, but their frequency  $h_{nm}$  decreases. To be general means few features and many entities, while to be special means many features and few entities.

**5:**

Overview of the different numbers:

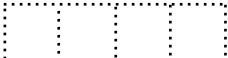


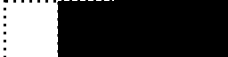
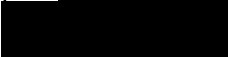


**6:**

*Measure* assign numbers to the quantities of their quality in order to be able to calculate them.

**7:**

Diversity:

min		$H = (1/4 \log_2 4/4 + 1/4 \log_2 4/4 + 1/4 \log_2 4/4 + 1/4 \log_2 4/4)$	= 0 bit,
medium		$H = (1/4 \log_2 4/3 + 1/4 \log_2 4/3 + 1/4 \log_2 4/3) + (1/4 \log_2 4/1)$	= 0,81 bit,
max		$H = (1/4 \log_2 4/2 + 1/4 \log_2 4/2) + (1/4 \log_2 4/2 + 1/4 \log_2 4/2)$	= 1 bit,
medium		$H = (1/4 \log_2 4/1) + (1/4 \log_2 4/3 + 1/4 \log_2 4/3 + 1/4 \log_2 4/3)$	= 0,81 bit,
min		$H = (1/4 \log_2 4/4 + 1/4 \log_2 4/4 + 1/4 \log_2 4/4 + 1/4 \log_2 4/4)$	= 0 bit.

**8:**

For continuous values, only the sums are replaced by integrals and the frequencies  $p_{nm}$  by their densities  $P_m$ :

$$H_m = \sum_n^{N_m} p_{nm} \log_2 1/p_{nm} \quad \rightarrow \quad H_m = \int_x^{X_m} P_m \log_2 1/P_m dx.$$

**9:**

$$p_{nm} = 1/N_m: \quad H_m = \sum_n^{N_m} p_{nm} \log_2 1/p_{nm} = \sum_n^{N_m} 1/N_m \log_2 N_m/1 = \log_2 N_m.$$

**10:**

Dependencies lead to changes, changes describe dependencies. Changes resolve contradictions and transform non-equilibrium into equilibrium states.

**11:**

Classical attractive or repulsive *forces* require spatial neighborhood: *locality*. Action at a distance takes place via indirect *transitive* effects or *quantum entanglement* instead of.

**12:**

The number of actual combinations may be limited not only because of dependence on X and Y ( $N_{XY} < N_X * N_Y$ ), but also because there are not enough entities in the given domain ( $n_{NM} < h$ ). Example: The number of license plates actually available is not limited solely by regulations, but because there are fewer cars than possible combinations.

This is another reason why changing the statistical distribution is a better way to measure information than simply reducing the number of samples.

**13:**

The tuples of the relation are implemented as columns, the sets as rows and the elements as identifiers. Tables do not contain all combinatorially possible rows, so they form subsets and thus describe the dependencies.

In the example of the Boolean table, all possible combinations of values of X and Y. This means that X and Y are independent of each other. The variable Z depends on X and Y. Its function here is  $Z = X \wedge Y$ . This reduces the eight possible rows to four. If the X-Y-table, without Z, is reduced from four to two lines, then X and Y become dependent on each other.  $Y = X$  for the red pair and  $Y = \neg X$  for the green pair.

X	Y	Z
0	0	0
0	1	1
1	0	1
1	1	1

**14:**

Information is the difference of diversity,  $I = \Delta H$ , both denote the same quantity with the unit of measurement "bit". Length and its change also represent a physical quantity and are both measured in "meter".

The different markings with "I" and "H" use familiar symbols and serve to better distinguish them.

**15:**

The information of an area can also be interpreted as actual information and the remaining diversity as potential information. The sum of both is the maximum diversity:

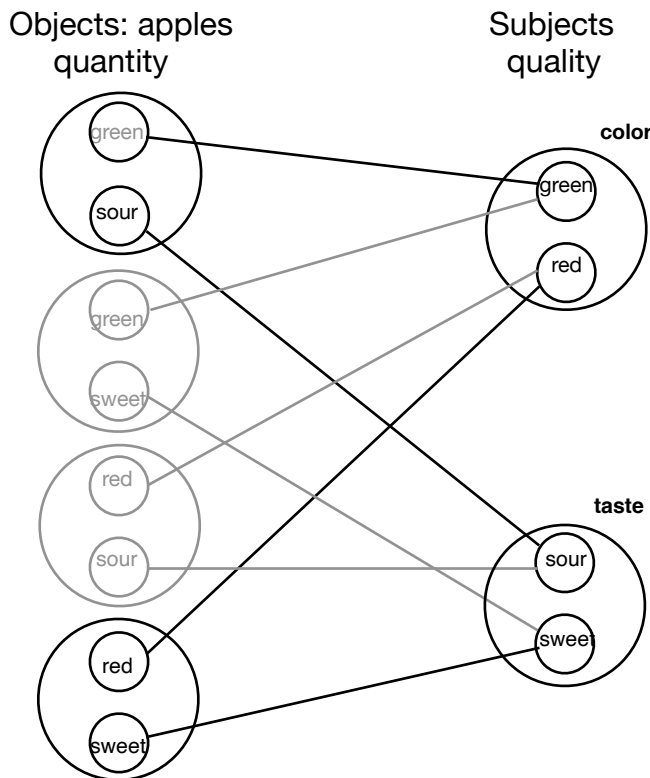
$$\begin{aligned}
 I_{\text{aktual}} &= H_{\text{maximum}} - H_{\text{potential}} \\
 H_{\text{maximum}} &= I_{\text{aktual}} + H_{\text{potential}} \\
 &= \text{Id} ( N_1 * N_2 * .. * N_M / N_{12..M} ) + \text{Id} N_{12..M} \\
 &= \text{Id} N_1 * N_2 * .. * N_M.
 \end{aligned}$$

**16:**

Instead of the terms quantity, quality, entity, dependency, object or subject, one could also use value, type, variable, signal, environment or system.

**17:**

Interdependence occurs between different qualities of an object. This means that certain combinations are forbidden, such as good and cheap or green and sweet:



**18:**

Subjects S are space areas; their environment is the remainder of the world,  $\neg S$ . They represent their environment by assigning their states to those of the environment.

Subjects are smaller than their environment, have fewer possible states than this, so:

$$N_S < N_{-S}.$$

Subjects are more complex than their environment, their internal information is greater than zero:

$$I_{int} > 0.$$

**19:**

Perception means that an object/environmental state changes a subject state. This makes the object active and the subject passive. The dichotomy of action/perceiving should therefore better mean acting/experiencing (/adversity, /suffering). Adaptation reduces these activities.

**20:**

Externally the quantities are distributed in physical space and can only be reached through movement. Internally they are spatially concentrated as qualities (senses) in state space. Externally, causality, as a direct effect, is only possible between spatial neighbors. However, only indirect effects, correlation, are possible between distant entities. Internally everything is adjacent, causality is possible.

States are stored internally, in that they are independent of each other, and are not changed.

**21:**

C.F.v.Weizsäcker (Weizsäcker 1985) speaks of "Kreislänge": Complex topics cannot be described and understood as a sequence, but only as a circle. A circle with no beginning or end can be started anywhere.

This text also uses terms that will only be defined later. It describes descriptions.

The principle of self-application was already applied by the author (Sösemann 1989) by "Programming of programming".

**22:**

The minimal, universal and complete programming language "Leibniz" (Sösemann 2010) served to search for the essence of languages and to introduce programming.

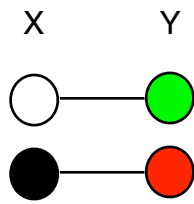
It is *turing-complete*. *Programming paradigms* such as structured, functional, logical or object-oriented can be defined and applied in it . The syntax is very short with three rules:

```
assignment = sequence [ ':' sequence ].
sequence   = symbol [ sequence ].
symbol     = '(' assignment ')'
           | '<' symbol '>'
           | '_'
           | sign.
```

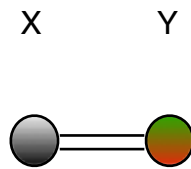
Semantics is formally defined by rules for parameterized, recursive word replacement. The assignment of symbol sequences describes interacting values. Sequences realize space or time. Recursion makes it possible to represent hierarchies.

**23:**

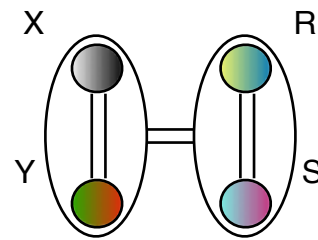
Hierarchical assignment graphs summarize the individual assignments of elements, as lines, to double lines, for the assignment of sets. This does not apply if there is complete assignment, i.e. independence. Borders separate the hierarchy levels:



element assignment



set assignment



hierarchical set assignment

**24:**

Colloquially, "information" relates to the information measure defined above as a set  $M = \{m_1, \dots, m_N\}$  relates to its cardinality  $N = |M|$ . The measure is the reduction of  $N_M$  to 1, i.e.  $\text{Id } N_M$ , the "information" is the specification of  $m_n$  from  $M$ .

**25:**

Example of true knowledge  $S(R)(Y(X))$  with variants 1 and 2:

X   Y	R   S	XY   RS	X Y R S	X Y R S
0   0	<del>0</del>   <del>0</del>	00   01	0-0   0 0	0-0   0 0
<del>0</del>   1	0   1	00   10	1-1   1 1	1-1   1 1
1   0	<del>1</del>   <del>0</del>	11   01		
1   1	<del>1</del>   <del>1</del>	11   10		

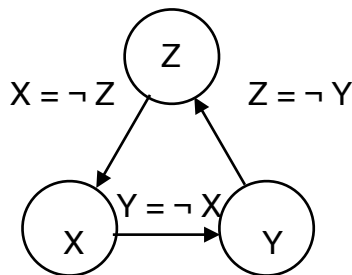
$$I_{XYRS} = \text{Id } 4/2 * 4/2 * 4/2 = \text{Id } 16/2 = 3,$$

$$Y = X, S = \neg R, X = S, Y = \neg R, \rightarrow Y = \neg R = S = X,$$

$$Y = X, S = \neg R, X = R, Y = \neg S, \rightarrow Y = \neg S = R = X.$$

**26:**

In some assignments, contradictions arise. Over time, these lead to *oscillations*, i.e. cyclical changes in the states involved. The double negation in the example results in identity. A third negation, instead of identity, causes all values to change cyclically:



time:	...	1	2	3	4	5	6	7	8	9	...
X:	...	0	1	0	1	...					
Y:	...	1	0	1	0	...					
Z:	...	0	1	0	1	...					

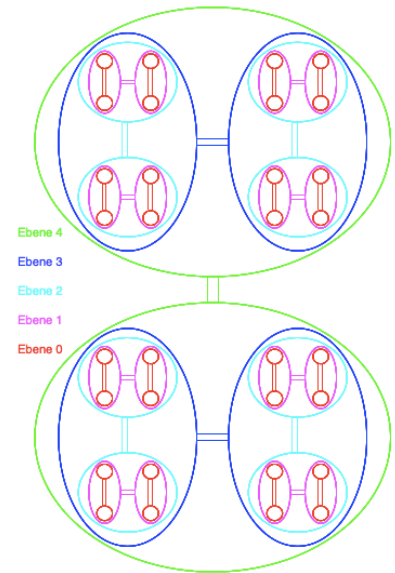
**27:**

Hawkins (*Hawkins 2006*) characterizes intelligence as the brain's ability to predict what will happen using analogies. Creativity is creating predictions by analogy. The future can be predicted by analogy from the past. The premise for this is that the world has structure and thus is predictable.



**28:**

- Level 4:** intelligence  
= knowledge  $\leftarrow$  thinking  $\rightarrow$  knowledge
- Level 3:** knowledge  
= information  $\leftarrow$  perceiving  $\rightarrow$  information
- Level 2:** information  
= state  $\leftarrow$  dependency  $\rightarrow$  state
- Level 1:** state  
= quality  $\leftarrow$  assignment  $\rightarrow$  quality
- Level 0:** quality  
= quantity  $\leftarrow$  exclusion  $\rightarrow$  quantity



**29:**

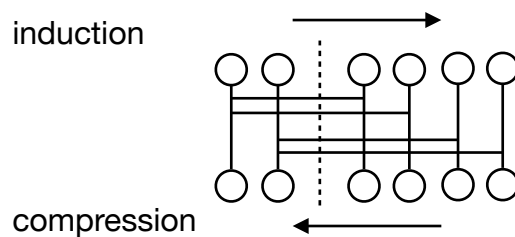
Constancy, i.e. the same quantity over long periods of time, means maximum temporal dependence. Spatially, however, the qualities must be independent of each other and not influence each other.

They are therefore not ergodic. This is what a system is called ergodic if the time means and the set means lead to the same result with probability one (*ergodic hypothesis*).

*Variables* Y are spatial entities, the quantities  $X_i$  (values), certain quality X (*type*) actively assigned from outside,  $Y_{t^*} := X_i$ , and kept until the next assignment,  $Y_t = X_i$ .

**30:**

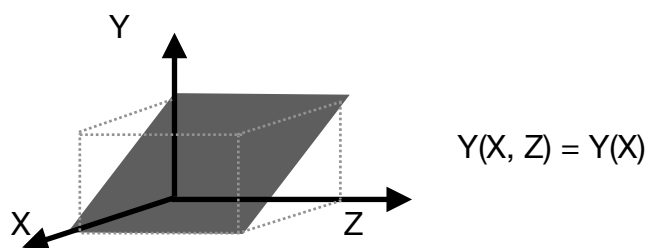
For the survival of organisms, it is often important to process knowledge faster than the environment changes. Compression replaces extensive knowledge with more compact knowledge, such as describing sequences of the same elements by repeating them.



**31:**

Redundancy can also be useful, such as in reducing disruptions in communication or as knowledge, when duplicating information reduces effort.

The dimension Z in the image is irrelevant for the permitted states, but X and Y are dependent on each other, so one of them is redundant:



Examples of redundancy and relevance of Boolean functions:

X Y I	$Z_i$	
-----		
0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1		$Z_1=0,$
0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 1		$Z_2=\neg X \wedge \neg Y,$
1 0 1 0 0 0 0 1 1 1 1 0 0 0 1 1 1		$Z_3=\neg X \wedge Y,$
1 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1		$Z_4=\neg X$
		$Z_5=X \wedge \neg Y,$
		$Z_6=\neg Y,$
		$Z_7=X \neg Y,$
		$Z_8=\neg X \wedge \neg Y,$
		$Z_9=X \wedge Y,$
		$Z_{10}=X \neg Y,$
		$Z_{11}=Y,$
		$Z_{12}=\neg X \vee Y,$
		$Z_{13}=X,$
		$Z_{14}=X \vee \neg Y,$
		$Z_{15}=X \vee Y,$
		$Z_{16}=1$
i = 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16		

In the functions  $Z_1$  and  $Z_{16}$ , X and Y are irrelevant and therefore redundant in  $Z_6$  and  $Z_{11}$  is X, and in  $Z_4$  and  $Z_{13}$ , Y is irrelevant and therefore redundant. In  $Z_7$  is  $Y=\neg X$  and in  $Z_{10}$  is  $Y=X$ , so one of each is redundant. In  $Z_8, Z_{12}, Z_{14}$  and  $Z_{15}$ , X and Y are relevant and partly redundant.

**32:**

Karl Friston is a neuroscientist and founder of the free energy theory, which experts believe could revolutionize brain research. Scientists believe that free energy theory allows for the creation of a grand unified theory of the brain, which can then be expressed in a mathematical law.

**33:**

In "The Markov Blanket Trick..." (*Raja 2021*), Vicente Raja et al criticize Friston's free energy principle: "We argue that FEP is not the general principle it is claimed to be, and that active inference is not the all-encompassing process theory it is purported to be either. The core aspects of our argumentation are that (i) FEP is just a way to generalize Bayesian inference to all domains by the use of a Markov blanket formalism, a generalization we call the Markov blanket trick; and that (ii) active inference presupposes successful perception and action instead of explaining them."

**34:**

"Perception is not a passive outside-in process—in which information is extracted from impressions on our sensory epithelia from “out there.” It is a constructive inside-out process—in which sensations are used to confirm or disconfirm hypotheses about how they were generated, a generative model."

**35:**

From (*Friston 2016*) and (*Friston 2022*): "Formal equivalence between the minimization of variational free energy and the maximization of model evidence in approximate Bayesian inference. Active Inference provides a novel first principle perspective to understand optimal behavior." Free energy as energy minus entropy, complexity minus accuracy and surprise plus a divergence term:

$F[Q, y]$ , as it is a functional (function of a function) of the approximate posterior  $Q$  and a function of data  $y$ :

$$\begin{aligned}
 F[Q, y] &= \underbrace{-\mathbb{E}_{Q(x)}[\ln P(y, x)]}_{\text{Energy}} - \underbrace{H[Q(x)]}_{\text{Entropy}} \\
 &= \underbrace{D_{KL}[Q(x) \parallel P(x)]}_{\text{Complexity}} - \underbrace{\mathbb{E}_{Q(x)}[\ln P(y|x)]}_{\text{Accuracy}} \\
 &= \underbrace{D_{KL}[Q(x) \parallel P(x|y)]}_{\text{Divergence}} - \underbrace{\ln P(y)}_{\text{Evidence}}
 \end{aligned}$$

## References

- Brillouin, Leon: Science and Information Theory. Academic Press 1956.
- Callen, H.B.: Thermodynamics and an Introduction to Thermostatistics. 2nd ed., John Wiley & Sons, New York 1985
- Friston, Karl: *The free-energy principle: a unified brain theory?* Nature Reviews 2010.
- Friston, Karl; u.a.: *Active inference and learning*. Neuroscience&Behavioral Reviews 68, 2016.
- Friston, K.; Parr, T.; Pezzulo, G.: *Active Inference - The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press 2022.
- Hawkins, Jeff: Die Zukunft der Intelligenz. rororo 2006.
- Junghans, Christian: Intentionalität. Begriffsdefinition in Hans Jörg Sandkühler (Hg.): Enzyklopädie Philosophie 1999.
- Raja, Vicente u.a.: *The Markov Blanket Trick: On the Scope of the Free Energy Principle and Active Inference*. ScienceDirekt 2021.
- Sösemann, Friedrich: Information, physikalische Entropie und Objektivität. Wiss.Zeitschrift der TH Karl-Marx-Stadt, Heft 1 1975.  
<https://friedrich-soesemann.de/data/documents/InfEntObj1975.pdf>.
- Sösemann, Friedrich: *Programmieren des Programmierens*. 1989.
- Sösemann, Friedrich: *Programmieren lernen mit "Leibniz"*. 2010.
- Weizsäcker, Carl Friedrich: Aufbau der Physik. Hanser 1985.

# Intelligentes Beschreiben und das Prinzip der Freien Energie

Friedrich Sösemann 5/2024

"Man muß die Dinge so einfach wie möglich machen.  
Aber nicht einfacher."

Albert Einstein

## Begriffe

*Qualität, Quantität, Zustand, Vielfalt, Häufigkeit, Abhängigkeit, Kausalität, Korrelation, Relation, Zuordnung, Information, Objekt, Subjekt, Beschreibung, Wahrheit, Wissen, Intelligenz, Prinzip der freien Energie.*

## Zusammenfassung

Information bemisst die Abhängigkeit von Zuständen, Wissen die zwischen Objekt- und Subjekt-Zuständen, Intelligenz die zwischen Subjekt-Zuständen. Beschreibungen speichern Objekt-Zustände.

Fristons Prinzip der freien Energie ist intelligent, verbindet Physik, Informatik und Biologie, ist aber nicht neu.

## Problem

Beschreibungen sind einfach, aber nicht zu einfach, wenn sie wahr und intelligent sind.

Im Folgenden werden, so einfach wie möglich, aber dennoch präzise, Beschreibungen selbst beschrieben. Sie und ihre Maße Information, Wissen und Intelligenz werden aus einer gegebenen Weltsicht abgeleitet.

Danach wird versucht Karl Fristons Prinzip der freien Energie einzuordnen und zu bewerten.

## Qualität und Quantität

Die *Welt* besteht aus unterschiedlichen *Quantitäten* im *Raum*. Die Quantitäten verändern sich in der *Zeit* im Rahmen ihrer *Qualität*.

Die Qualität und Quantität wird mathematisch beschrieben durch *Mengen*, deren Elemente sich gegenseitig ausschliessen.

Beispiele dazu sind *Eigenschaften* und deren Werte, *Alphabete* und deren Buchstaben oder *Zahlensysteme* und deren Ziffern /1/.

*Kontinuierliche* Qualitäten besitzen unendlich viele, geordnete Quantitäten, *diskrete* zumeist endlich viele, die auch ungeordnet sein können (*Nominalskalen*).

Diskret sind die natürlichen Zahlen, kontinuierlich die reellen, Die Eigenschaft Farbe, mit der Werten rot, grün oder blau, ist diskret und nominal, als *Spektrum* hingegen kontinuierlich und *ordinal*.

*Entitäten* sind Orte oder Bereiche aus mehreren benachbarten Orten. Ihnen sind gewöhnlich mehrere Qualitäten zugeordnet.

Die Gesamtheit der Quantitäten aller vorhandenen oder betrachteten /2/ Qualitäten werden als *Zustand* bezeichnet und mathematisch als *Tupel* dargestellt.

Beispiele dafür sind Zahlen, *Worte* oder die Darstellung von Zuständen als Punkte im *Zustands-Raum* und in *Tabellen*.

## Häufigkeit und Vielfalt

Die Grösse des gewählten Bereiches und seine Qualitäten bestimmen die Zahl der Entitäten. Diese und die Anzahl /3/ ihrer Quantitäten bestimmen den Umfang möglicher Zustände.

Die Anzahl der möglichen unterschiedlichen Zustände  $N_M$  ergibt sich als Produkt der Zahlen der Quantitäten  $N_m$  aller  $M$  betrachteten Qualitäten:

$$N_M = \prod_m^M N_m.$$

Die Anzahl der tatsächlich vorhandenen Zustände, also die Zahl der Entitäten  $H_M$ , berechnet sich aus der Zahl möglicher Zustände und deren Häufigkeit  $h_{nm}$  /4/:

$$H_M = \sum_m^M \sum_n^{N_m} h_{nm}.$$

Die *absolute Häufigkeit*  $h_{nm}$ , dividiert durch die Zahl der Entitäten  $h_m$  der Qualität  $m$  ergibt deren *relative Häufigkeit* oder *statistische Wahrscheinlichkeit*  $p_{nm}$  /5/:

$$p_{nm} = h_{nm} / h_m.$$

Die Anzahl von Quantitäten steht auch für die Vielfalt einer Qualität. Ein Maß /6/ der Vielfalt  $H$  sollte für mehrere Qualitäten additiv sein, dabei aber berücksichtigen, dass die Anzahl multiplikativ wächst, und bei nur einem möglichen Zustand den Wert Null ergeben soll. Das erfüllt der *Logarithmus* der Anzahl von Zuständen. Seine Basis Zwei ergibt dann die Maßeinheit Bit:

$$H_M = \sum_m^M H_m = \text{ld} \prod_m^M h_m.$$

Eine Qualität mit gleich häufigem Auftreten seiner Quantitäten wird als vielfältiger wahrgenommen, als bei vorwiegenden Auftreten der einen und nur seltenem der anderen /7/.

Um das auch im Maß  $H_m$  widerzuspiegeln, wird die relative Häufigkeit  $p_{nm}$  des Vorkommens der Entitäten berücksichtigt und deren Summe entsprechend gewichtet /8/:

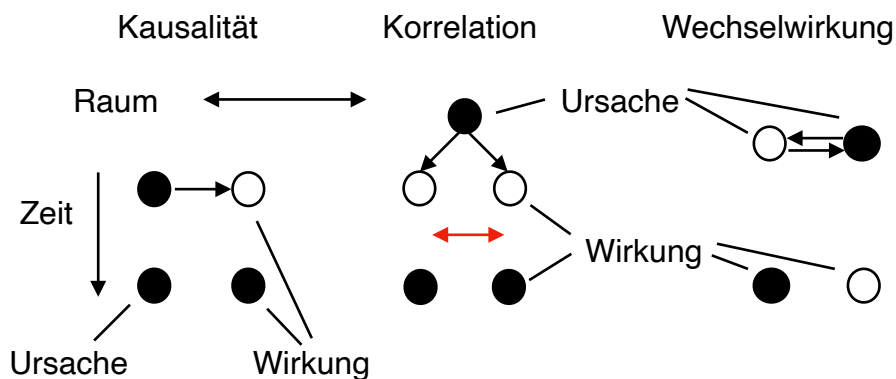
$$H_m = \sum_n^{N_m} p_{nm} \text{ld } 1/p_{nm}.$$

Bei Gleichverteilung wird die Vielfalt maximal und geht in  $H_m = \text{ld } N_m$  über /9/.

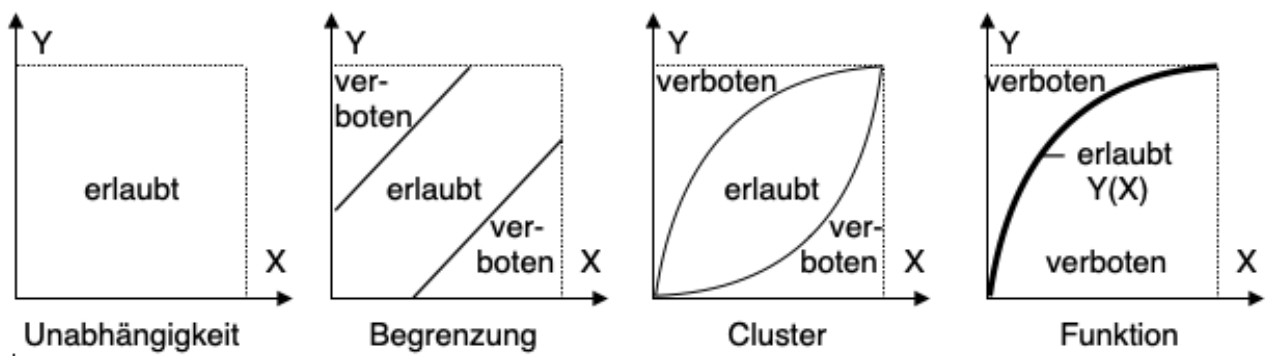
## Abhängigkeit und Information

Qualitäten wirken aufeinander: Die Quantitäten ihrer Qualität schliessen einander aus und die Quantitäten unterschiedlicher Qualitäten hängen voneinander ab, bevorzugen oder hemmen sich gegenseitig. So erzeugen sie *Korrelation*, *Kausalität* oder *Wechselwirkung*.

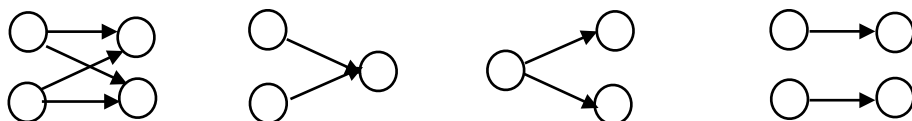
Die Kausalität ist *asymmetrisch*, unterscheidet Ursache von Wirkung. Die Korrelation ist *symmetrisch* und wird durch Kausalität von dritter Seite verursacht. Die symmetrische Wechselwirkung ist gegenseitige Kausalität.



Die Abhängigkeit /10/ kann als Anziehen oder Abstoßen /11/, als erlaubte oder verbotene Zuordnungen von Quantitäten /12/ und in Form natürlicher Gesetze oder künstlicher Regeln und Definitionen auftreten:



Bildlich kann die Abhängigkeit durch Graphen:



Zuordnung: vollständig unabhängig, eindeutig unscharf, mehrdeutig zufällig, eineindeutig determiniert

mathematisch durch *Relationen*:

$$X = \{ x_1, \dots, x_{N_x} \}, Y = \{ y_1, \dots, y_{N_y} \}, R(X, Y) \subseteq X \otimes Y,$$

das sind Untermengen des Kreuzproduktes von Mengen, oder praktisch durch *Tabellen* /13/ dargestellt werden.

Die *Information* /14/ ist das Maß der Abhängigkeit von Zuständen, die Differenz der abhängigen von der unabhängigen, maximalen Vielfalt /15/:

$$I = H_{\max} - H = \text{Id } N_{\max} / N.$$

Das ist bei der Auswahl einer von N ausschliessenden Quantitäten:

$$I = \text{Id } N,$$

bei M voneinander abhängigen Qualitäten über die Anzahlen:

$$I = \text{Id } N_1 * N_2 * \dots * N_M / N_{12..M}$$

oder detaillierter über die Häufigkeiten:

$$I = \sum_{n_1}^{N_1} \sum_{n_2}^{N_2} \dots \sum_{n_M}^{N_M} p_{n_1 n_2 \dots n_M} * \text{Id} ( p_{n_1 n_2 \dots n_M} / p_{n_1} * p_{n_2} * \dots * p_{n_M} ).$$

## Subjekt und Beschreibung

*Objekte* sind Entitäten deren Zustand relativ konstant ist und sich von dem ihrer Umgebung unterscheiden. *Subjekte* sind Objekte, die vergangene Zustände speichern können /16/.

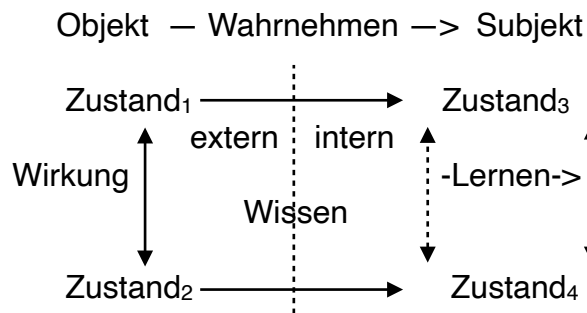
Objekte sind passiv, reagieren, werden behandelt und wahrgenommen. Subjekte sind aktiv, agieren, handeln und nehmen wahr /17/.

*Beschreibungen* sind passiv, wie Objekte, speichern, wie Subjekte und bilden externe in internen Zuständen ab.

Alle drei beinhalten Information, müssen also wenigstens zwei verschiedene Quantitäten besitzen /18/. Subjekte und Beschreibungen realisieren *Intentionalität* (Junghans 1999), weisen auf andere Entitäten.

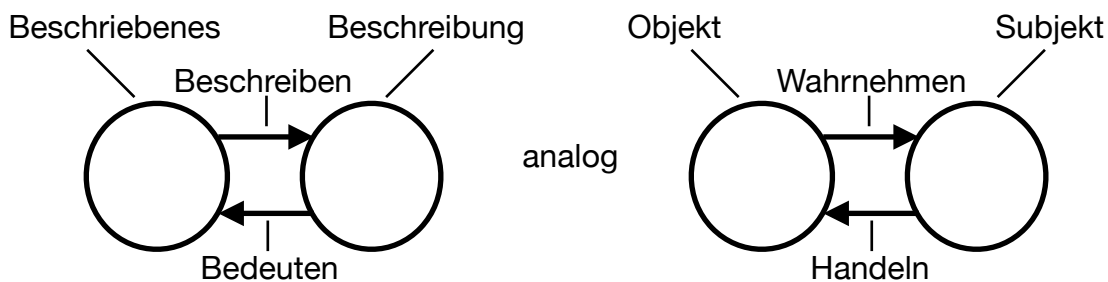
Objekte werden von Subjekten wahrgenommen indem externe Zustände internen zugeordnet werden /19/. Die Abhängigkeit der externen Zustände begrenzt somit transitiv auch die internen.

Lernen speichert dann die wahrgenommene Kombinationen, ergänzt so indirekte durch direkte Zuordnungen, ersetzt Korrelation durch Kausalität /20/:



Die gelernten redundanten Zuordnungen erhöhen die Effizienz von Subjekten, einmal gelernte Zuordnungen müssen nicht mehrfach aus der Umgebung ermittelt werden.

Beschreibungen sind Objekte, welche die beschriebenen Objekte vertreten. Ihre Zustände bilden passiv externe Zustände ab und speichern sie, verwandeln Zeit in Raum.

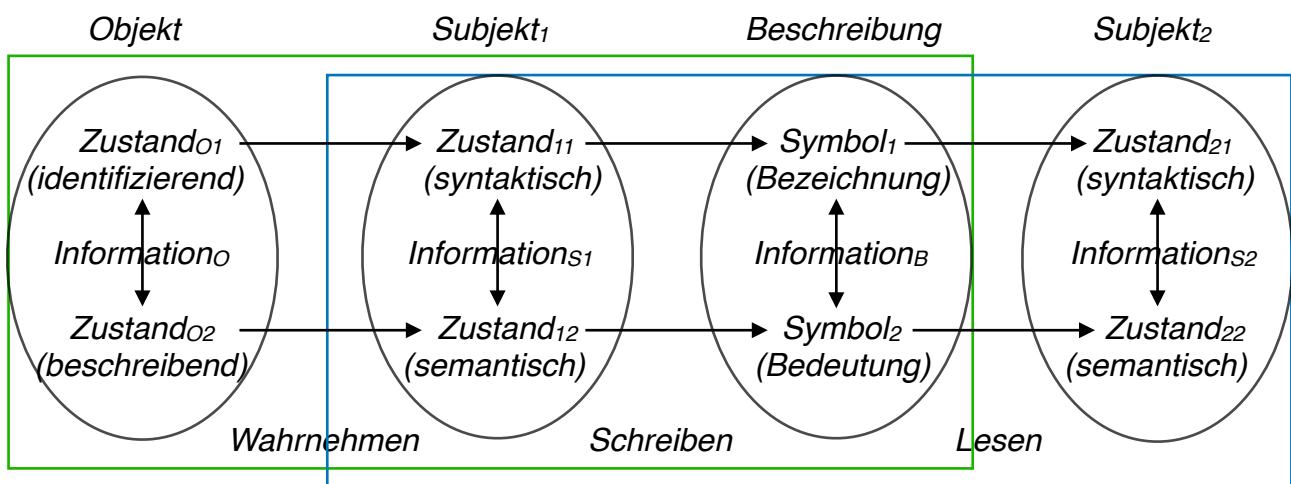


Beschreibungen sind zumeist Folgen von Symbolen, welche über ihre Bezeichnungen einander zugeordnet sind.

Hören und Lesen, Sprechen und Schreiben sind sequentiell, das Beschriebene aber zumeist ein Netz von Entitäten und Beziehungen. Darum bahnt sich die Beschreibung als Pfad durch das Netz, gibt es *rekursive Definitionen*, *Top-Down* oder *Bottom-Up* Vorgehen /21/.

Fehler, Widerspruch, Vollständigkeit oder Wahrscheinlichkeit sind dabei keine Merkmale der Realität, also des Beschriebenen, sondern der Beschreibung.

Beschreibungen dienen dem Speichern von Wissen und dessen Austausch zwischen Subjekten:





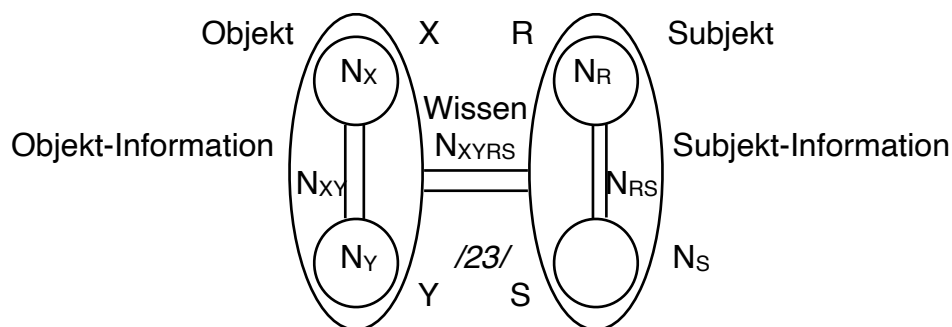
Beschreibungen sind Zustände einer *Sprache*. Deren *Syntax*, *Semantik* und *Pragmatik* sind intersubjektiv vereinbarte Definitionen /22/. Dabei verkörpert die Syntax, die Struktur, ihre Information, die Semantik, die Bedeutung, ihr Wissen und die Pragmatik, die Effizienz, ihre Intelligenz:

Ebene	Begriff	Maß	Beschreibung
0	Wert, Wirkung	$N_1, N_2, N_{12}$	Wörter
1	Information	$I_{XY} = \text{Id}(N_X * N_Y / N_{YX})$	Syntax
2	Wissen	$I_{XYRS} = \text{Id}(N_{XY} * N_{RS} / N_{XYRS})$	Semantik
3	Intelligenz	$I_M = \text{Id}(N_1 * N_2 * .. * N_M / N_{12..M})$	Pragmatik

## Wahrheit und Wissen

Die Abhängigkeit zwischen zwei Bereichen mit abhängigen Zuständen, wie Objekt und Subjekt, kann als *Wissen* interpretiert werden, ist Information über Informationen:

$$I_{OS} = \text{Id} ( N_{XY} * N_{RS} / N_{XYRS} ).$$



Das Wissen zwischen den Bereichen plus den internen Informationen der Bereiche ergibt die Information des Gesamt-Bereiches:

$$\begin{aligned} I_{XYRS} &= I_{XY} + I_{RS} + I_{(XY)(RS)} \\ &= \text{Id} ( N_X * N_Y / N_{XY} ) * ( N_R * N_S / N_{RS} ) * ( N_{XY} * N_{RS} / N_{XYRS} ) \\ &= \text{Id} N_X * N_Y * N_R * N_S / N_{XYRS}. \end{aligned}$$

Information als Maß von Abhängigkeit darf nicht mit dem Inhalt von Informationen, der umgangssprachlichen "Information", verwechselt werden /24/. So ist auch Wissen zwar notwendig, aber nur *wahres* Wissen ist hinreichend.

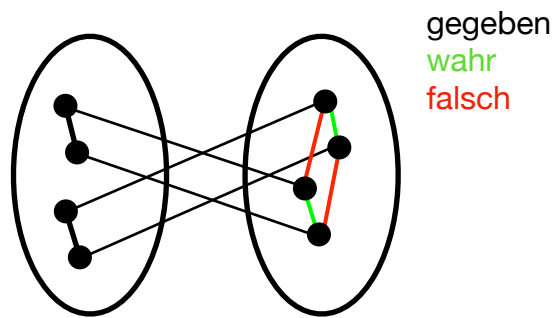
Subjekte besitzen wahres Wissen als internes Modell von Objekten oder ihrer Umgebung. Modelle von Subjekten S(R) sind bezüglich des Objektes Y(X) wahr, wenn für ihre Funktionen

$$Y( S( R( X ) ) ) = Y( X )$$

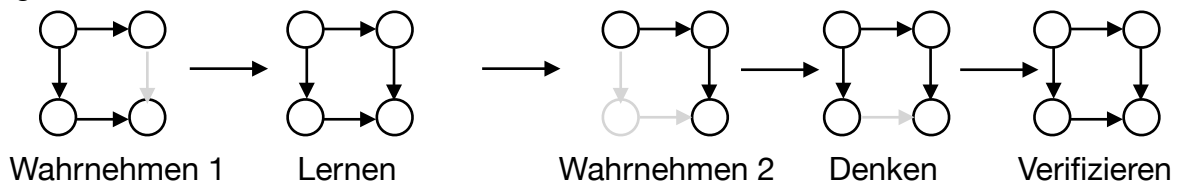
gilt, wobei R(X) und S(Y) die Wahrnehmungen, X(R) und Y(S) das Handeln von Subjekten sind /25/.

Wahrheit erfordert strukturerhaltende Zuordnungen, *Isomorphie*:

$$I_{\text{Wahr}} = I_d / 2 = I_{\text{Falsch}}$$



Die Wahrheit von Wissen wird *verifiziert*, wenn die vorausgesagte Quantität tatsächlich vorliegt:



Wissen setzt kein Bewusstsein voraus. So ist beim Thermostat die externe Information - heizen erhöht die Temperatur - mit der internen Information - wenn die Temperatur zu niedrig, dann mehr heizen - fest verknüpft.

Bei Beschreibungen gibt es vier Mitspieler: das Objekt oder die Umwelt  $O$ , das schreibende Subjekt  $S_1$ , die Beschreibung  $B$  und das lesende Subjekt  $S_2$ . Das führt zu drei verbundenen Wissens-Beziehungen,  $I_{OS_1}$ ,  $I_{S_1B}$  und  $I_{BS_2}$ , die *transitiv* das gewünschte Wissen  $I_{OS_2}$  ergeben:

$$I_{OS_2} = I_{OS_1} + I_{S_1B} + I_{BS_2} = I_d N_O * N_{S_1} * N_B * N_{S_2} / N_{OS_1BS_2}$$

Sind alle drei Teilwissen wahr, so ist es auch das transitive Gesamtwissen. Bei Fiktionen fehlt das Objekt, bei Fehlern, Missverständnissen oder Lügen ist wenigstens eine Zuordnung falsch.

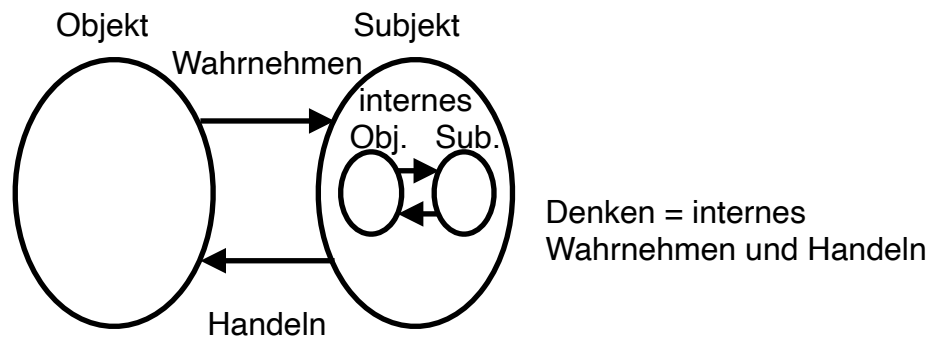
Widersprüchliches Wissen ( $X=x_i$  &  $X=x_j$  mit  $i \neq j$ ) in Beschreibungen bildet keine externe Information ab (*disjunkte* Quantitäten). Widersprüche zwischen Objekten, Beschreibungen und Subjekten können korrigiert werden, *dialektische* Widersprüche in Objekten oder der Umwelt führen zu Zyklen /26/.

Präzise Beschreibungen ordnen Zustände eineindeutig zu, vermeiden so falsche Verallgemeinerungen, Mißverständnisse oder Nachfragen.

Beschreibungen sind *verständlich*, wenn Beschreibungs- auf Leser-Zustände abbildbar sind.

## Aufwand und Intelligenz

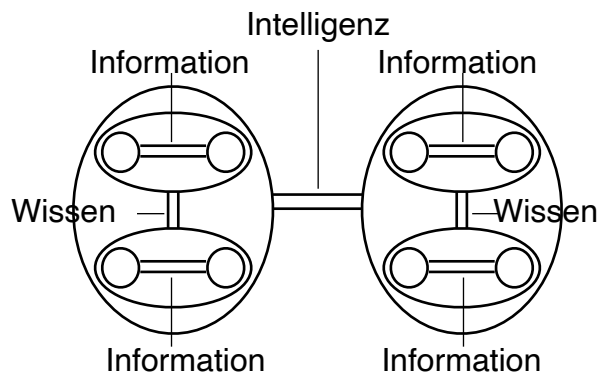
Lernen schafft Wissen, Denken nutzt Wissen, ersetzt externes durch internes Navigieren. Dadurch agiert das Subjekt effizienter und sicherer als direkt in der Umwelt /27/.



Intelligenz ist das Maß der Effizienz des Denkens, durch Verringerung des Aufwandes. Das wird erreicht, indem auch die Abhängigkeiten zwischen den Wissens-Elementen gelernt werden. Die interne Struktur wird verbessert, indem die Wissens-Menge in ein Wissens-Netz verwandelt wird.

*Intelligenz* ist das Maß der Abhängigkeit von Wissen, der Reduktion der Zahl von Verbindungsmöglichkeiten aller Wissens-elemente:

$$I_{XYRSM} = \text{Id} ( N_{XYRS1} * N_{XYRS2} * .. * N_{XYRSM} / N_{XYRS12..M} ).$$



Intelligenz ist damit eine weitere Ebene /28/ von Abhängigkeiten und erfordert eine zusätzliche Raum-Dimension:

Information	1 Dimension		2 Qualitäten
Wissen	2 Dimensionen		Meta-Information 4 Qualitäten
Intelligenz	3 Dimensionen		Meta-Wissen 8 Qualitäten

Beispiel für die drei Abhängigkeits-Maße:

Information	$l_{o1} = \{ (x_1, y_1), (\overset{x}{x}_1, y_2), (\overset{x}{x}_2, y_1), (x_2, y_2) \}$	$= \text{Id } 4/2 = 1 \text{ Bit,}$
	$l_{o2} = \{ (\overset{x}{x}_1, y_1), (x_1, y_2), (x_2, y_1), (\overset{x}{x}_2, y_2) \}$	$.. = 1 \text{ Bit,}$
	$l_{s1} = \{ (r_1, s_1), (\overset{r}{r}_1, s_2), (\overset{r}{r}_2, s_1), (r_2, s_2) \}$	$.. = 1 \text{ Bit,}$
	$l_{s2} = \{ (\overset{r}{r}_1, s_1), (r_1, s_2), (r_2, s_1), (\overset{r}{r}_2, s_2) \}$	$.. = 1 \text{ Bit,}$

Wissen	$l_{OS1} = \{ (l_{O1}, l_{S1}), (\cancel{l_{O1}}, \cancel{l_{S2}}), (\cancel{l_{O2}}, \cancel{l_{S1}}), (l_{O2}, l_{S2}) \}$	..	= 1 Bit,
	$l_{OS2} = \{ (\cancel{l_{O1}}, \cancel{l_{S1}}), (l_{O1}, l_{S2}), (l_{O2}, l_{S1}), (\cancel{l_{O2}}, \cancel{l_{S2}}) \}$	..	= 1 Bit,
Intelligenz	$l_{OS12} = \{ (l_{OS1}, l_{OS1}), (\cancel{l_{OS1}}, \cancel{l_{OS2}}), (\cancel{l_{OS2}}, \cancel{l_{OS1}}), (l_{OS2}, l_{OS2}) \}$	..	= 1 Bit.

Das *Speichern* vergangener Zustände unterscheidet Subjekte von Objekten. Das Denken, als Nutzen und Verarbeiten von Wissen, setzt voraus, dass die internen Zustände, welche das Wissen verkörpern, über die Zeit erhalten bleiben.

Speicher verwandeln Zeit in Raum, sind Entitäten mit unabhängigen konstanten Zuständen /29/.

*Denken*, wie *Deduktion*, *Induktion* oder *Kompression*, ist Wissensverarbeitung. Deduktion schließt aus implizitem auf explizites Wissen. Induktion erzeugt neues Wissen aus vorhandenem. Induktion vermehrt, Kompression /30/ verringert die Anzahl von Wissens-Zuständen im Subjekt.

Der Aufwand, die *Arbeit*, als Kosten zum Speichern und Verarbeiten von Information sollen möglichst gering sein. Er hängt von der Menge und Art, sowie von der *Relevanz* und *Redundanz* /31/ des gespeicherten Wissens ab.

Auch Beschreibungen speichern Wissen. Intelligente Beschreibungen sind relevant, redundanzarm und explizit:

Relevante Beschreibungen betreffen und verbessern den Zustand des Lesers, sind oft Ergebnis einer Suche und Auswahl der Beschreibung selbst.

Mehrfaches Lesen der gleichen Beschreibung oder mehrfache, identische Aussagen in einer Beschreibung ergeben Redundanz. Sie kann durch einfaches Auftreten und Angabe der Zahl ihrer Wiederholungen vermieden werden. Redundanz kann aber auch Fehler bei der Übermittlung vermeiden oder aber das Lernen verfestigen, "Wiederholung ist die Mutter des Studiums".

*Explizite* Beschreibungen ersparen dem Leser aus *implizitem* Wissen zu deduzieren.

Ist das Prinzip minimaler freier Energie nun eine intelligente Beschreibung?

## Prinzip der freien Energie

Das *Prinzip der freien Energie* von Karl Friston will Leben, Geist und KI definieren und wird gegenwärtig diskutiert, befürwortet /32/ und kritisiert /33/.

Grundlage sind Organismen die ihre lebensnotwendigen Zustände in einer ständig sich ändernden Umwelt erhalten müssen.

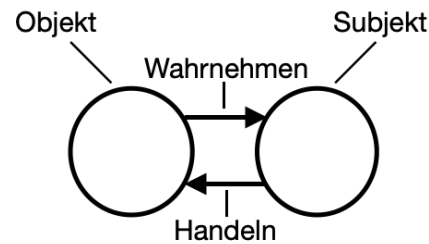
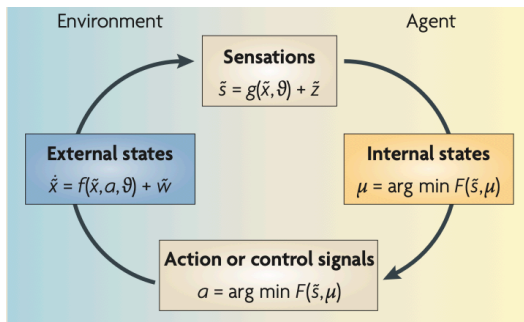
Dazu machten Gehirne ständig Vorhersagen über die Welt /34/ und ziehen *Bayessche Schlüsse*, um Überraschungen so weit wie möglich zu vermeiden. Denn Energie wird gespart, indem *Aufmerksamkeit* nur dann aktiviert wird, wenn etwas Neues oder Ungewöhnliches passiert.

Friston wendet nun das bereits existierende *Minimumsprinzip der freien Energie* auf die Schnittstelle zwischen Organismen und ihre Umwelt an.

Das Prinzip besagt, weil die Gesamtenergie  $E$  konstant bleibt, die gebundene  $E_{\text{geb}}$ , gemäß dem 2. Hauptsatz der Wärmelehre, einem Maximum zustrebt, die freie Energie  $E_{\text{frei}}$  einem Minimum zustreben muss:

$$\begin{array}{l} \text{Gesamtenergie} = \text{Wärme} + \text{Arbeit} \\ \text{konstant} \quad \quad \quad \text{maximal} \quad \text{minimal} \\ E \quad \quad \quad = \quad E_{\text{geb}} \quad + \quad E_{\text{frei}}. \end{array}$$

Dabei strukturiert er die Wechselwirkung zwischen dem Subjekt und seiner Umgebung (siehe Bild links aus (Friston 2010)) wie oben (siehe rechts):



Er betrachtet Bayessches Schliessen, Lernen durch Vorhersage und minimale freie Energie als zusammen gehörig, jedoch:

1. Bayessches Lernen bedingt eine statistische Sichtweise, andere Formen sind aber möglich (siehe oben).
2. Zum Lernen durch Vorhersagen ist bereits ein internes Umweltmodell erforderlich. Allerdings kann das schrittweise durch Vorhersagen mit eventuellen Korrekturen, gewonnen werden. Doch nur wenn Umlernen effizienter ist als Neulernen wird dabei der Lernaufwand verringert.
3. Der Begriff der *freien Energie*  $F$  wurde schon 1882 von Helmholtz geprägt. Mit  $U$  als interne Energie,  $T$  als Temperatur und  $S$  als Entropie gilt:

$$F = U - T \cdot S.$$

Auch das Minimumsprinzip der freien Energie ist schon länger bekannt (Callen 1985). Die Verbindung von Information und physikalischer Entropie wurde schon 1956 von Leon Brillouin beschrieben (Brillouin 1956) und vom Autor 1975 diskutiert (Sösemann 1975). Sein "Prinzip der minimalen freien Energie" benutzt nun Karl Friston (Friston 2010) um Leben, Geist und Künstliche Intelligenz zu definieren:

$$\begin{array}{l} \text{Energie} = \text{Wärme} + \text{Arbeit} \\ E \quad \quad = \quad E_{\text{geb}} \quad + \quad E_{\text{frei}}. \end{array}$$

Er weist aber auch auf andere Interpretationen hin /35/ und verknüpft es mit Bayesscher Schlussweise, der Informationstheorie und Kognition:

Statistical physics	Bayesian inference and information theory	Cognitive interpretation
Minimize variational free energy	Maximize model evidence (or marginal likelihood); minimize surprisal (or self-information)	Perception and action
Minimize expected free energy; Hamiltonian principle of least Action	Infer the most likely (or less surprising) course of action	Planning as inference
Attain nonequilibrium steady-state	Perform approximate Bayesian inference	Self-evidencing
Gradient flows on energy functions; gradient descent on free energy	Gradient ascent on model evidence; gradient descent on surprisal	Neuronal dynamics

Angemessener, als über Energie, scheint die Erklärung über freie und gebundene Vielfalt, die Information:

$$\begin{aligned} \text{Vielfalt} &= \text{gebundene Vielfalt} + \text{freie Vielfalt} \\ \text{Id } N_X * N_Y &= \text{Id } N_X * N_Y / N_{XY} + \text{Id } N_{XY}. \end{aligned}$$

Das Muster ist gleich, aber die Begriffe passender. Das Minimieren freier Energie ist Beseitigen der Fehler des internen Modells.

Ein Subjekt mit optimalem und wahren Welt-Wissen braucht nicht mehr zu lernen oder die Umgebung nicht mehr zu verändern, ist mit der Umwelt im Gleichgewicht. Über die Objekt/Subjekt-Grenze sind dann keine Wirkungen mehr erforderlich. Der minimalen freie Energie entspricht maximales Wissen.

Subjekte unterscheiden sich von Objekten oder ihrer Umwelt durch höhere Komplexität, der Möglichkeit zu speichern um die Welt auch in der Vergangenheit abzubilden, und sie korrelieren stärker als nur wechselwirkende Objekte, kurz: sie besitzen wahres Wissen.

Ein *Isomorphismus* zwischen Wissensgebieten weist auf eine übergeordnete Struktur oder eine *Hierarchie* auf einander aufbauender Schichten.

*Analogie*, sollte fruchtbar sein, Strukturen und Gesetze des mehr erschlossenen Gebietes auf das andere übertragen werden.

Bei Hierarchie müsste die Biologie aus der Physik abgeleitet werden, unter Berücksichtigung von *Emergenz* in höheren Schichten.

Von beidem ist Friston's Prinzip noch entfernt, kann aber als Anregung, verkrustete Sichten aufzubrechen, gesehen werden.

## Anmerkungen

Die Anmerkungen sind aus dem Haupttext verlagert, um das Lesen dort flüssiger zu gestalten. Dabei erleichtern *kursive* Links Hin- und Rücksprünge. Andere *kursive* Textstellen verweisen auf externe Quellen, wie Wikipedia-Einträge oder Online-Artikel.

### 1:

Raum und Zeit sind keine Qualitäten, sondern Behälter oder Anordnung der Quantitäten. In der absoluten Raumauffassung von *Newton* ist der Raum ein Behälter. In der relationalen Auffassung von *Leibniz* spannen die Quantitäten den Raum erst auf. So sind Orte abgeleitete, redundante Quantitäten, Orte ohne Quantität gibt es nicht.

### 2:

Ob die Qualitäten und Quantitäten, und damit die Entitäten, *objektiv* existieren oder *subjektiv* konstruiert werden ist die philosophische Frage, ob *Naturalismus* oder *Konstruktivismus* gültig ist. Das zu beantworten ist hier nicht vorgesehen.

### 3:

Gezählt werden können Objekte/Ereignisse eines Raum/Zeit-Bereiches, die Elemente einer Menge, die Häufigkeit der Elemente von Multimengen oder die Eigenschaften eines Objektes in Produktmengen:

$$\begin{aligned}
 | ( 0, 1, 1, 0, 0, 0, 1, 1 ) | &= 8 && \text{Entitäten in Gesamtheit,} \\
 | \{ 0, 1, 1, 0, 0, 0, 1, 1 \} | = | \{ 0, 1 \} | &= 2 && \text{Werte in Menge,} \\
 | ( 0, 1, 1, 0, 0, 0, 1, 1 ) | = | \{ (0, 4), (1, 4) \} | &= 2 && \text{Häufigkeiten in Multimenge,} \\
 | \{ (0, 1), (1, 0), (0, 0), (1, 1) \} | = | \{ 0, 1 \} \times \{ 1, 0 \} | &= 4 && \text{Zustände in Produktmenge.}
 \end{aligned}$$

Problematisch zu zählen sind kontinuierliche Größen. Wo werden die Grenzen zwischen den zu zählenden Elementen gezogen? Intersubjektive Vereinbarungen, wie beispielsweise die Definition der Frequenzbereiche von Farben, oder eine gegebene Messgenauigkeit gehen ja in die Anzahl ein.

### 4:

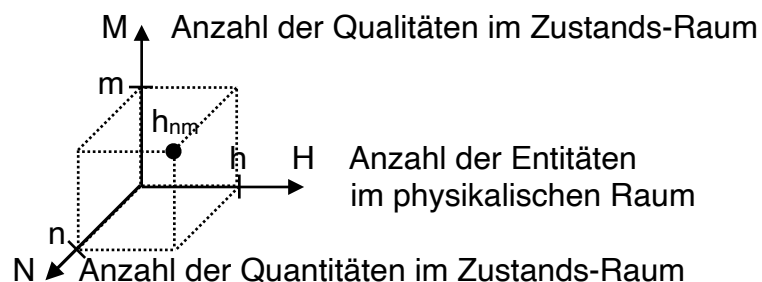
Damit Häufigkeiten  $h_{nm} > 1$  auftreten, muss die Zahl der Entitäten grösser als die Zahl unterschiedlicher Zustände sein:  $H_M > N_M$ .

Zum Finden aller auftretenden Quantitäten der betrachteten Qualitäten muss ein grosser Raum, die "Welt", über eine lange Zeit durchsucht werden. In zu kleinen Bereichen treten mögliche Zustände dann nicht auf:  $h_{nm}=0$ .

Wird die Zahl berücksichtigter Qualitäten  $M$  erhöht, so steigt die Zahl möglicher Zustände  $h_{NM}$ , aber deren Häufigkeit  $h_{nm}$  nimmt ab. Allgemein sein heisst wenig Merkmale und viele Entitäten, Speziell sein hingegen viele Merkmale und wenig Entitäten.

### 5:

Überblick zu den verschiedenen Anzahlen:



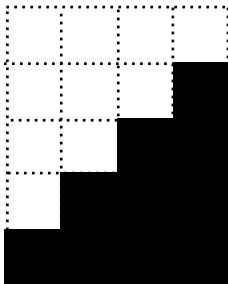
**6:**

Maße ordnen den Quantitäten ihrer Qualität Zahlen zu, um damit rechnen zu können.

**7:**

Vielfalt:

min



$$H = (1/4 \log_2 4/4 + 1/4 \log_2 4/4 + 1/4 \log_2 4/4 + 1/4 \log_2 4/4) = 0 \text{ Bit,}$$

mittel

$$H = (1/4 \log_2 4/3 + 1/4 \log_2 4/3 + 1/4 \log_2 4/3) + (1/4 \log_2 4/1) = 0,81 \text{ Bit,}$$

max

$$H = (1/4 \log_2 4/2 + 1/4 \log_2 4/2) + (1/4 \log_2 4/2 + 1/4 \log_2 4/2) = 1 \text{ Bit,}$$

mittel

$$H = (1/4 \log_2 4/1) + (1/4 \log_2 4/3 + 1/4 \log_2 4/3 + 1/4 \log_2 4/3) = 0,81 \text{ Bit,}$$

min

$$H = (1/4 \log_2 4/4 + 1/4 \log_2 4/4 + 1/4 \log_2 4/4 + 1/4 \log_2 4/4) = 0 \text{ Bit.}$$

**8:**

Für kontinuierliche Werte werden lediglich die Summen durch Integrale und die Häufigkeiten  $p_{nm}$  durch ihre Dichten  $P_m$  ersetzt:

$$H_m = \sum_n^{N_m} p_{nm} \log_2 1/p_{nm} \quad \rightarrow \quad H_m = \int_x^{X_m} P_m \log_2 1/P_m dx.$$

**9:**

$$p_{nm} = 1/N_m: \quad H_m = \sum_n^{N_m} p_{nm} \log_2 1/p_{nm} = \sum_n^{N_m} 1/N_m \log_2 N_m/1 = \log_2 N_m.$$

**10:**

Abhängigkeiten führen zu Veränderungen, Veränderungen beschreiben Abhängigkeiten. Veränderungen lösen Widersprüche, führen Nicht-Gleichgewichts- in Gleichgewichts-Zustände über.

**11:**

Klassische anziehende oder abstoßende *Kräfte* bedingen räumliche Nachbarschaft: *Lokalität*. Fernwirkung findet über indirekte *transitive* Wirkungen oder *Quantenverschränkung* statt.

**12:**

Die Anzahl tatsächlicher Kombinationen kann nicht nur wegen Abhängigkeit von X und Y eingeschränkt sein ( $N_{XY} < N_X * N_Y$ ), sondern auch weil es im gegebenen Bereich nicht genügend Entitäten gibt ( $h_{NM} < h$ ).

So sind die tatsächlich vorhandenen Auto-Kennzeichen nicht allein durch Vorgaben eingeschränkt, sondern weil es weniger Autos als Kombinationsmöglichkeiten gibt.

Auch darum ist die Veränderung der statistischen Verteilung besser als die bloße Verringerung von Anzahlen geeignet, Information zu bemessen.

**13:**

Die Tupel der Relation werden als Spalten, die Mengen als Zeilen und die Elemente als Bezeichner realisiert. Tabellen enthalten nicht alle kombinatorisch möglichen Zeilen, bilden dadurch Untermengen und beschreiben so die Abhängigkeiten.



Im Beispiel der Booleschen Tabelle treten alle möglichen Kombinationen der Werte von X und Y auf. Damit sind X und Y voneinander unabhängig. Die Variable Z ist von X und Y abhängig. Ihre Funktion ist hier  $Z = X \wedge Y$ . Das reduziert die acht möglichen auf vier Zeilen. Wird die X-Y-Tabelle, ohne Z, von vier auf zwei Zeilen reduziert, so werden X und Y voneinander abhängig. Für das rote Paar gilt dann  $Y = X$  und für das grüne  $Y = \neg X$ .

X	Y	Z
0	0	0
0	1	1
1	0	1
1	1	1

**14:**

Information ist die Differenz von Vielfalt,  $I = \Delta H$ , beide bezeichnen die gleiche Größe mit der Maßeinheit "Bit". Auch Länge und ihre Veränderung stehen für eine physikalische Größe und werden beide in "Meter" gemessen. Die verschiedene Kennzeichnung durch "I" und "H" verwendet gewohnte Symbole und dient der besseren Unterscheidung.

**15:**

Die Information eines Bereiches kann auch als aktuelle Information und die verbleibende Vielfalt als potentielle Information interpretiert werden. Beider Summe ist die maximale Vielfalt:

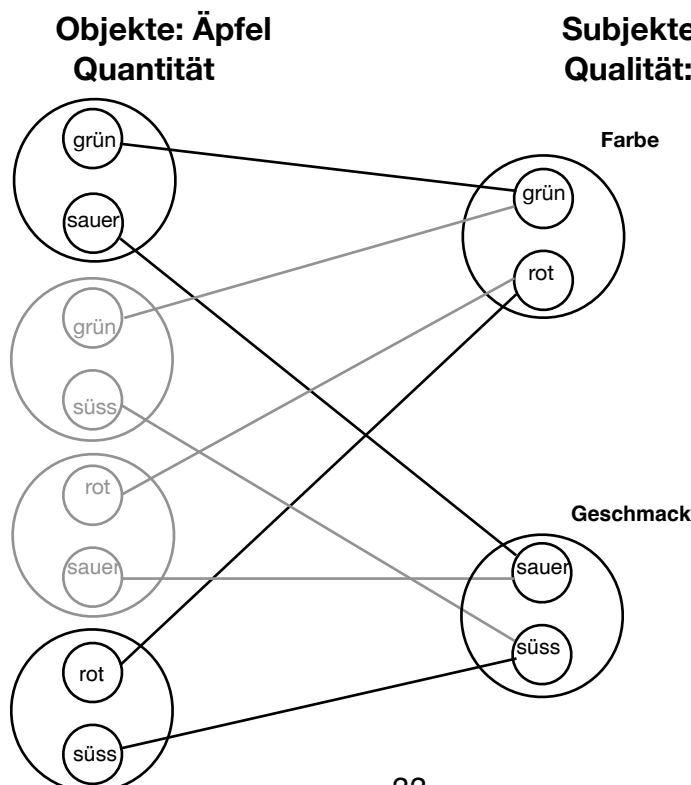
$$\begin{aligned}
 I_{\text{aktuell}} &= H_{\text{maximal}} - H_{\text{potentiell}} \\
 H_{\text{maximal}} &= I_{\text{aktuell}} + H_{\text{potentiell}} \\
 &= \text{Id} ( N_1 * N_2 * .. * N_M / N_{12..M} ) + \text{Id} N_{12..M} \\
 &= \text{Id} N_1 * N_2 * .. * N_M.
 \end{aligned}$$

**16:**

Statt der Bezeichnungen Quantität, Qualität, Entität, Abhängigkeit, Objekt oder Subjekt könnte auch Wert, Typ, Variable, Signal, Umgebung oder System verwendet werden.

**17:**

Gegenseitige Abhängigkeit tritt zwischen verschiedenen Qualitäten eines Objektes auf. Dadurch sind gewisse Kombinationen verboten, wie gut und billig oder grün und süßs:



**18:**

Subjekte S sind Raum-Bereiche; ihre Umwelt ist der verbleibende Rest zur Welt,  $\neg S$ . Sie bilden ihre Umwelt ab, indem ihre Zustände denen der Umwelt zugeordnet sind.

Subjekte sind kleiner als ihre Umwelt, besitzen weniger mögliche Zustände als diese, also:

$$N_s < N_{\neg s}.$$

Subjekte sind komplexer als ihre Umgebung, ihre interne Information ist grösser Null:

$$I_{int} > 0.$$

**19:**

Wahrnehmen bedeutet, dass ein Objekt/Umwelt-Zustand einen Subjekt-Zustand verändert. Damit ist aber das Objekt aktiv, das Subjekt passiv. Die Dichotomie Handeln/Wahrnehmen sollte also besser Handeln/Erfahren (/Widerfahren, /Erleiden) heissen. Anpassung verringert diese Aktivitäten..

**20:**

Extern sind die Quantitäten im physikalischen Raum verteilt und nur durch Bewegung erreichbar. Intern sind sie als Qualitäten (Sinne) im Zustands-Raum räumlich konzentriert. Extern ist Kausalität, als direkte Wirkung, nur zwischen räumlichen Nachbarn möglich. Zwischen entfernten Entitäten ist hingegen nur indirekte Wirkung, Korrelation, möglich. Intern ist alles benachbart, Kausalität möglich.

Zustände werden intern gespeichert, indem sie voneinander unabhängig sind und nicht verändert werden.

**21:**

C.F.v.Weizsäcker (Weizsäcker 1985) spricht von *Kreisgang*: Komplexe Themen können nicht als Folge, sondern nur als Kreis beschrieben und verstanden werden. Einen Kreis, ohne Anfang und Ende kann man an beliebiger Stelle beginnen.

Auch der vorliegende Text benutzt schon Begriffe, die er erst später definiert werden, beschreibt Beschreibungen. Das Prinzip der Selbstanwendung wurde vom Autor (Sösemann 1989) schon beim "Programmieren des Programmierens" angewendet.

**22:**

Die minimale, universelle und vollständige Programmiersprache "Leibniz" (Sösemann 2010) diene der Suche nach dem Wesenskern von Sprachen und der Einführung in das Programmieren.

Sie ist *turing-vollständig*. In ihr können *Programmierparadigmen*, wie strukturiert, funktional, logisch oder objektorientiert, definiert und angewendet werden. Die Syntax ist mit drei Regeln sehr klein:

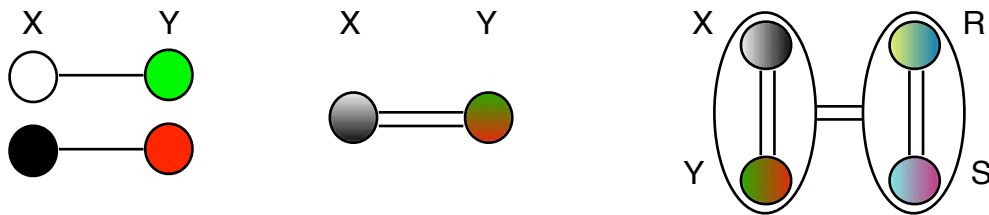
zuordnung	=	folge [ ':' folge ] .
folge	=	symbol [ folge ] .
symbol	=	'(' zuordnung ')'
		'<' symbol '>'
		'_'   zeichen .

Die Semantik wird formal durch Regeln zur parametrisierten, rekursiven Wortersetzung definiert.

Die Zuordnung von Symbolfolgen beschreibt wechselwirkende Werte. Folgen realisieren Raum oder Zeit. Die Rekursion ermöglicht Hierarchien darzustellen.

**23:**

Hierarchische Zuordnungs-Graphen fassen die einzelnen Zuordnungen von Elementen, als Striche, zum Doppelstrich, für die Zuordnung von Mengen, zusammen. Dieser entfällt bei vollständiger Zuordnung, also Unabhängigkeit. Umrandungen trennen dabei die Hierarchie-Ebenen:



Elemente-Zuordnung    Mengen-Zuordnung    Hierarchische Mengen-Zuordnung

**24:**

Umgangssprachliche "Information" verhält sich zum oben definierten Informations-Maß wie eine Menge  $M = \{ m_1, \dots, m_N \}$  zu deren Mächtigkeit  $N = |M|$ . Das Maß ist die Reduktion von  $N_M$  auf 1, also  $\text{Id } N_M$ , die "Information" ist die Angabe des  $m_n$  aus  $M$ .

**25:**

Beispiel wahren Wissens  $S(R)( Y(X) )$  mit Varianten 1 und 2:

X   Y	R   S	XY   RS	X Y R S	X Y R S
--- ---	--- ---	----- -----	0-0   0 0	0-0   0 0
0   0	0   0	00   01	0-0   0 0	0-0   0 0
0   1	0   1	00   10	1-1   1 1	1-1   1 1
1   0	1   0	11   01	1-1   1 1	1-1   1 1
1   1	1   1	11   10		

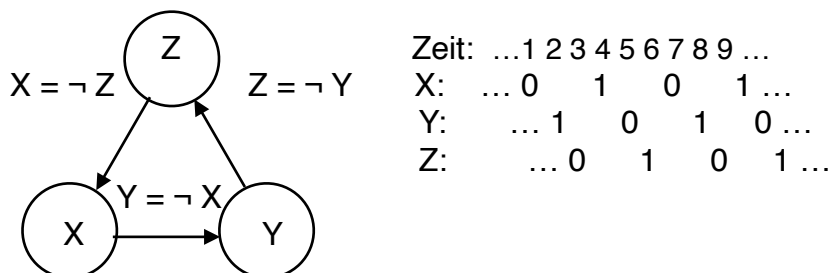
$$I_{XYRS} = \text{Id } 4/2 * 4/2 * 4/2 = \text{Id } 16/2 = 3,$$

$$Y = X, S = \neg R, X = S, Y = \neg R, \rightarrow Y = \neg R = S = X,$$

$$Y = X, S = \neg R, X = R, Y = \neg S, \rightarrow Y = \neg S = R = X.$$

**26:**

Bei manchen Zuordnungen treten Widersprüche auf. Diese führen im Zeitablauf zu *Oszillationen*, also zyklischen Veränderungen der beteiligten Zustände. Die doppelte Negation im Beispiel ergibt Identität. Eine dritte Negation, statt der Identität, läßt alle Werte zyklisch wechseln:



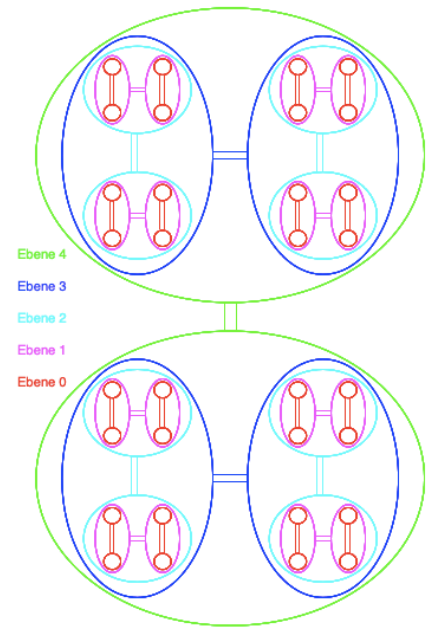
**27:**

Hawkins (*Hawkins 2006*) charakterisiert Intelligenz als die Fähigkeit des Gehirns, mit Hilfe von Analogien vorauszusagen, was geschehen wird. Kreativität ist das Erstellen von

Vorhersagen mittels Analogie. Zukunft kann per Analogie aus der Vergangenheit vorhergesagt werden. Prämisse dafür ist, dass die Welt Struktur besitzt und damit vorhersagbar ist.

**28:**

- Ebene 0:** Qualität  
= Quantität <—Ausschluss—> Quantität
- Ebene 1:** Zustand  
= Qualität <—Zuordnung—> Qualität
- Ebene 2:** Information  
= Zustand <—Abhängigkeit—> Zustand
- Ebene 3:** Wissen  
= Information <—Wahrnehmen—> Information
- Ebene 4:** Intelligenz  
= Wissen <—Denken—> Wissen



**29:**

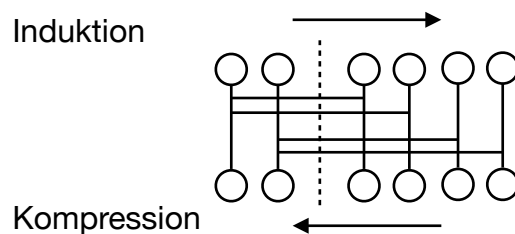
Konstanz, also gleiche Quantität über lange Zeitabschnitte, bedeutet maximale zeitliche Abhängigkeit. Räumlich müssen die Qualitäten hingegen voneinander unabhängig sein, sich gegenseitig nicht beeinflussen.

Sie sind damit nicht ergodisch. So wird ein System dann genannt, wenn die Zeitmittel und Scharmittel mit der Wahrscheinlichkeit eins zum gleichen Ergebnis führen (*Ergodenhypothese*).

*Variablen* Y sind räumliche Entitäten, die Quantitäten  $X_i$  (Werte), bestimmter Qualität X (*Typ*) aktiv von aussen zugewiesen bekommen,  $Y_t := X_i$ , und bis zur nächsten Zuweisung behalten,  $Y_t = X_i$ .

**30:**

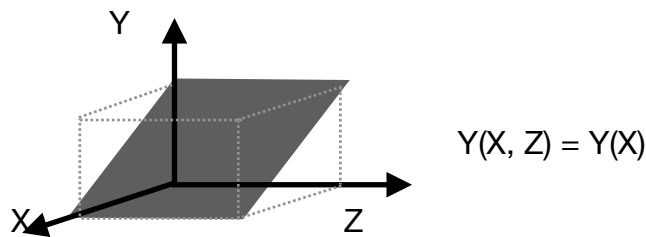
Für das Überleben von Organismen ist es oft wichtig, Wissen schneller zu verarbeiten als die Umgebung sich verändert. Kompression ersetzt umfängliches Wissen durch kompakteres, wie das Beschreiben von Folgen gleicher Elemente durch deren Wiederholung.



**31:**

Redundanz kann auch nützlich sein, wie beim Verringern von Störungen bei der Kommunikation oder bei Wissen, wenn das Duplizieren der Information den Aufwand reduziert.

Die Dimension Z im Bild ist für die erlaubten Zustände irrelevant, hingegen sind X und Y voneinander abhängig, also eines davon redundant:



Beispiele der Redundanz und Relevanz Boolescher Funktionen:

X	Y	I	$Z_i$																$Z_1=0,$	$Z_2=\neg X \wedge \neg Y,$
-----			-----																$Z_3=\neg X \wedge Y,$	$Z_4=\neg X$
0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	$Z_5=X \wedge \neg Y,$	$Z_6=\neg Y,$
0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	$Z_7=X \neg Y,$	$Z_8=\neg X \wedge \neg Y,$
1	0	1	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	1	$Z_9=X \wedge Y,$	$Z_{10}=X \neg Y,$
1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	$Z_{11}=Y,$	$Z_{12}=\neg X \vee Y,$
																			$Z_{13}=X,$	$Z_{14}=X \vee \neg Y,$
																			$Z_{15}=X \vee Y,$	$Z_{16}=1$
i =	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16				

In den Funktionen  $Z_1$  und  $Z_{16}$  ist X und Y, in  $Z_6$  und  $Z_{11}$  ist X und in  $Z_4$  und  $Z_{13}$  ist Y irrelevant und damit zugleich redundant. in  $Z_7$  gilt  $Y=\neg X$ , und  $Z_{10}$  gilt  $Y=X$ , also ist je eines redundant, in  $Z_8, Z_{12}, Z_{14}$  und  $Z_{15}$  sind X und Y relevant und teilweise redundant.

**32:**

Karl Friston ist Neurowissenschaftler und Begründer der Theorie der freien Energie, von der Experten glauben, dass sie die Gehirnforschung revolutionieren könnte. Wissenschaftler gehen davon aus, dass die Theorie der freien Energie die Schaffung einer großen einheitlichen Theorie des Gehirns ermöglicht, die dann in einem mathematischen Gesetz ausgedrückt werden kann.

**33:**

In "The Markov Blanket Trick ..." kritisieren Vicente Raja u.a. (Raja 2021) Fristons Prinzip der freien Energie:

"We argue that FEP is not the general principle it is claimed to be, and that active inference is not the all-encompassing process theory it is purported to be either. The core aspects of our argumentation are that (i) FEP is just a way to generalize Bayesian inference to all domains by the use of a Markov blanket formalism, a generalization we call the Markov blanket trick; and that (ii) active inference presupposes successful perception and action instead of explaining them."

**34:**

"Perception is not a passive outside-in process—in which information is extracted from impressions on our sensory epithelia from "out there." It is a constructive inside-out process—in which sensations are used to confirm or disconfirm hypotheses about how they were generated, a generative model."

### 35:

Aus (Friston 2016) und (Friston 2022):

"Formal equivalence between the minimization of variational free energy and the maximization of model evidence in approximate Bayesian inference. Active Inference provides a novel first principle perspective to understand optimal behavior."

Free energy as energy minus entropy, complexity minus accuracy and surprise plus a divergence term:

$F[Q, \gamma]$ , as it is a functional (function of a function) of the approximate posterior  $Q$  and a function of data  $\gamma$ :

$$\begin{aligned} F[Q, \gamma] &= \underbrace{-\mathbb{E}_{Q(x)}[\ln P(\gamma, x)]}_{\text{Energy}} - \underbrace{H[Q(x)]}_{\text{Entropy}} \\ &= \underbrace{D_{KL}[Q(x) \parallel P(x)]}_{\text{Complexity}} - \underbrace{\mathbb{E}_{Q(x)}[\ln P(\gamma | x)]}_{\text{Accuracy}} \\ &= \underbrace{D_{KL}[Q(x) \parallel P(x | \gamma)]}_{\text{Divergence}} - \underbrace{\ln P(\gamma)}_{\text{Evidence}} \end{aligned}$$

## Literatur

- Brillouin, Leon: Science and Information Theory. Academic Press 1956.
- Callen, H.B.: Thermodynamics and an Introduction to Thermostatistics. 2nd ed., John Wiley & Sons, New York 1985
- Friston, Karl: *The free-energy principle: a unified brain theory?* Nature Reviews 2010.
- Friston, Karl; u.a.: *Active inference and learning*. Neuroscience&Behavioral Reviews 68, 2016.
- Friston, K.; Parr, T.; Pezzulo, G.: *Active Inference - The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press 2022.
- Hawkins, Jeff: Die Zukunft der Intelligenz. rororo 2006.
- Junghans, Christian: Intentionalität. Begriffsdefinition in Hans Jörg Sandkühler (Hg.): Enzyklopädie Philosophie 1999.
- Raja, Vicente u.a.: *The Markov Blanket Trick: On the Scope of the Free Energy Principle and Active Inference*. ScienceDirekt 2021.
- Sösemann, Friedrich: Information, physikalische Entropie und Objektivität. Wiss.Zeitschrift der TH Karl-Marx-Stadt, Heft 1 1975.  
<https://friedrich-soesemann.de/data/documents/InfEntObj1975.pdf>.
- Sösemann, Friedrich: *Programmieren des Programmierens*. 1989.
- Sösemann, Friedrich: *Programmieren lernen mit "Leibniz"*. 2010.
- Weizsäcker, Carl Friedrich: Aufbau der Physik. Hanser 1985.