

# Information, knowledge and intelligence as a hierarchy of relations

*Friedrich Sösemann 6/2023*

## Terms

*World, Relation, Information, Complexity, Perception, Learning, Knowledge, Thinking, Intelligence, Describing.* (link to remark /0/)

## Abstract

Information, knowledge and intelligence are defined as a hierarchy of relations: Information as dependent properties, knowledge as dependent information, and intelligence as dependent knowledge. The same dependency measure applies to all three.

Syntax, semantics and pragmatics of descriptions embody information, knowledge and intelligence.

The precision and measurability of these terms should reduce vagueness and contradictions in their application.

## Approach

*Information* is on everyone's lips. Although formulated by *Shannon* more than seventy years ago as removed indeterminacy, it is still interpreted very differently. This complicates the overview and mutual exchange. For *knowledge* and *intelligence*, there is a complete lack of uniform definitions and measures.

In order to solve this problem, a precise and general definition of these terms is sought here. Precision and generality are to be achieved through the use of abstract mathematical means.

But this approach leads to new problems:

The derivation of the terms as a hierarchy of relations requires the use of description, truth and adequacy. Applying them even before they are defined is a vicious circle.

The way out is offered by recursive definitions: from the simplest term, those are derived that itself presupposes. This closes the circle.

The price of generality through abstraction is the loss of detail. But only insignificant features may be abstracted from. For example, information and knowledge are perceived as directed, but their measure is undirected.

Exact and universal mathematical terms alone are not enough. They lack meaning, their assignment to the world. The world must first be defined by an *ontology* that is as minimal and universal as possible.

## World

The *world* consists of *values* /1/ in space and time and the values *affect* each other.

*Space* is the ordered juxtaposition of places, *time* is the ordered and directed succession of times (points in time) /2/.

Multiple values can occur in one place at a time. These are assigned to various *properties* (*dimensions*). The values of a property never occur together in one place at one time: the apple is red or green, never red and green, but it is red at the top and green at the bottom, or green now and red later. Each location has a value for each property at all times: The apple is red, sweet, round and expensive.

The totality of all values of a place or of all places of a spatial area at a time is called *state*.

If the number of places, areas or times is greater than that of different values, they occur several times, have different *frequencies*.

The effect /3/ between directly adjacent values prevents or favors their *combination*, or allows them to occur more rarely or more frequently. Neighborhood is made possible through the ordering of space and time.

Only values of different properties in one place at a time, such as red and sweet, or adjacent values of the same property, such as red in the front and green in the back, are possible.

The *dependency* creates change /4/, but also *structures, laws and rules* /5/.

## Relation

The mathematical *relation* defines the dependency of two sets of values as a subset of their product:

$$X = \{ x_1, \dots, x_{N_x} \}, Y = \{ y_1, \dots, y_{N_y} \}, R(X,Y) \subseteq X \otimes Y. \quad /6/$$

The sets represent properties, their elements are the values and the product creates all possible combinations. The subset contains only the combinations remaining because of the dependency.

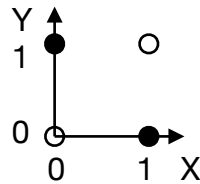
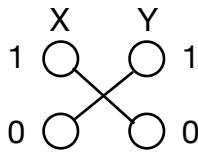
If the properties are independent, all combinations are possible, the size of the relation is then the product of the size of the initial sets:

$$|R| = |X| * |Y|.$$

In the case of complete dependency, on the other hand, only one-to-one assignments occur:

$$|R| = |X| = |Y|.$$

Relations are also represented by graphs, diagrams or tables. The example shows the one-to-one relation of the negation of binary quantities  $X = \neg Y$ :



X	Y
0	1
1	0

## Information

Information is defined as the dependence of properties.

The dependency can be directional, as in "information about", or non-directional, as in structures. But their measure is undirected /7/ and thus leaves all possibilities open.

$$I_{XY} = \text{ld} ( N_X \cdot N_Y / N_{XY} ), \quad N_X \begin{array}{c} \text{X} \\ \circ \end{array} \begin{array}{c} \text{Y} \\ \circ \end{array} N_Y \quad /8/$$

be the measure of the dependency between the properties X and Y, with  $N_X \cdot N_Y$  as the size of the product set  $X \otimes Y$ , and  $N_{XY}$  as the size of the subset  $R(X, Y)$ . Their quotient describes the dependency: if this increases, the number of actual combinations in the denominator decreases and the value of the quotient increases.

So that the information is zero in the case of independence and other properties can be added /9/, although the numbers are multiplying, the logarithm of the quotient becomes used /10/. The binary logarithm "ld" leads to the unit of measure "bit".

In the example of the negation of binary properties, the four possible assignments are reduced to two. This gives  $I_{XY} = \text{ld} 2^2/2 = \text{ld} 2 = 1$  bit.

The information differs from other correlation measures only in the definition of its measure /11/.

The information of all values of one property, the maximum or self-information, results with  $N_X = N_Y = N_{XY} = N$ :

$$I_X = \text{ld} ( N \cdot N / N ) = \text{ld} N. \quad /12/$$

Dependence can not only reduce the number of associated values, it can also change their frequency. Then the frequency of the values  $p_n$  is used instead of the number N. The frequency of the actual joint occurrence  $p_{xy}$  is set in relation to the independence  $p_x \cdot p_y$ . The information measure of two unequally distributed properties is then the weighted mean over all combinations:

$$I_{XY} = \sum_x \sum_y \frac{N_x N_y}{N} p_{xy} \cdot \text{ld} ( p_{xy} / p_x \cdot p_y ).$$

Information is also defined for continuous properties. The discrete frequencies  $p_{xy}$  are replaced by continuous densities  $P(x, y)$  and the sums by integrals:

$$I_{XY} = \int \int P(x, y) \cdot \text{ld} ( P(x, y) / P(x) P(y) ) dx dy.$$

## Complexity

The information of more than two properties of an area is also called *complexity*. The variety of proposed measures is even greater for her than for information.

Their definition only requires the expansion of two quantities to M (as in /9/):

$$I_M = \text{Id} ( N_1 * N_2 * .. * N_M / N_{12..M} ).$$

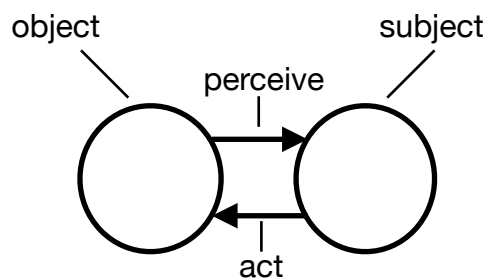
The information within an area is called "internal", outside "external", or also "actual" and "potential". The sum of both is the maximum information:

$$\begin{aligned} I_{\max} &= I_{\text{int}} + I_{\text{ext}} = I_{\text{akt}} + I_{\text{pot}} \\ &= \text{Id} ( N_1 * N_2 * .. * N_M / N_{12..M} ) + \text{Id} N_{12..M} \\ &= \text{Id} N_1 * N_2 * .. * N_M. \end{aligned}$$

In the negation example, the following applies:  $I_{\text{int}} = \text{Id} 2^*2/2 = 1 \text{ Bit}$ ,  $I_{\text{ext}} = \text{Id} 2 = 1 \text{ Bit}$ ,  
 $I_{\max} = \text{Id} 2^*2 = 2 \text{ Bit}$ .

## Perception

Perception is the directed dependence between an *object* and a *subject*, or between a subject and its environment. *Perception* and *acting* create dependency for both:



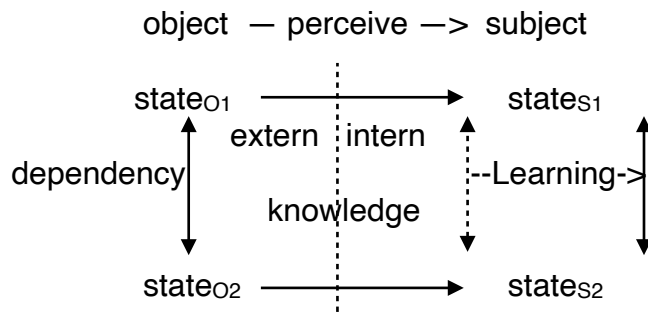
Values are perceived. They are combined into properties in the subject and are roughly specified by different senses such as seeing, hearing and feeling.

If Action and perception alternate, targeted learning is enabled, but also the external production of desired objects, so-called *artifacts*.

Subjects perceive only part of the values of their environment. Their complexity is necessarily less than that of the world. This then causes the *chance* by ignorance /13/.

## Learning

Learning creates knowledge: If not only individual values, but several and their presence or absence are perceived, the states of the object are recorded and mapped to states of the subject, the sensations:



By moving the subject in space or time, it perceives, stores and connects several values. Learning stores external information internally, replacing transitive correlation with causation.

*Storage* transforms succession into juxtaposition, transforms time into space.

Objects and subjects differ in the variability of their relations. This is constant for objects, initially variable and constant for subjects through learning. The relations of the subjects adapt to those of the objects /14/.

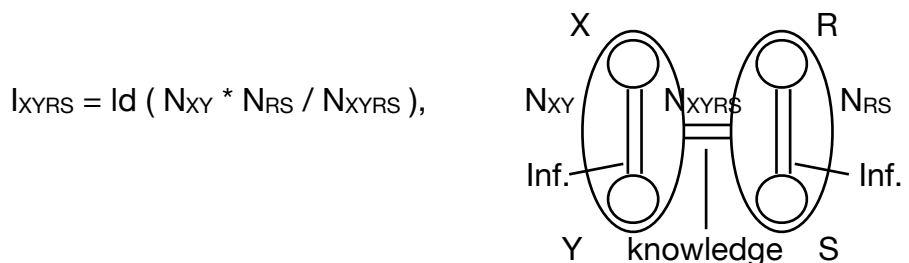
The example shows how the fixed negation relation of the object is realized via the identities of perception in the initially independent and variable subject:

object		perception		subject	
X	Y	X	R	Y	S
0	1	0	0	0	0
1	0	1	1	1	1
				0	1
				1	0
				<del>1</del>	<del>1</del>

## Knowledge

Knowledge is defined as the dependence on information.

Perceived and acting knowledge are directed in opposite directions. However, its measure, the dependency between object and subject, is undirected:



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

with  $N_{XY}$  as the number of external and  $N_{RS}$  that of internal information;  $N_{XYRS}$  is the number of actual assignments of both.

The sum of the complexity of object, subject and knowledge results in the overall complexity:

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

# Thinking

Thinking links knowledge and thus creates intelligence.

Thinking forms properties from values, laws and rules from states, structures from elements and *conclusions* from assumptions.

Examples are: red and green are colors; 01 and 10, but not 00 and 11, define negation; apples are green and sour or red and sweet;  $a < b$  and  $b < c$  gives  $a < c$ .

Perceived knowledge is elementary, consists of separate facts, the mapping of external to internal information.

Simple facts can be found faster in external physical space than compound ones /15/. Facts are thus preferably perceived and learned. Only in the subject are they united into theories through thinking.

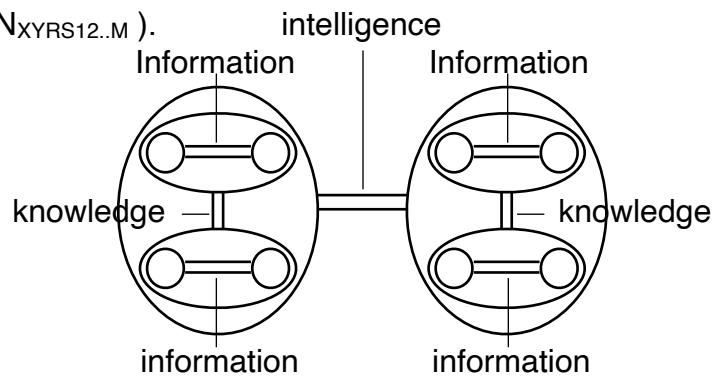
In physical space, only values in the same place at the same time or directly adjacent can affect each other (*locality*). In the subject, on the other hand, all values are adjacent, because of all neurons can be connected to one another via *axons*, and *feature space* replaces physical space.

# Intelligence

*Intelligence* is defined as the dependence on knowledge, its complexity.

It and its measure are undirected. The measure, is the *complexity* of knowledge, is the reduction in the number of possible connections between knowledge elements:

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



level 1:	information	=	state	<-effect-->	state
level 2:	knowledge	=	information	<-perception-->	information
level 3:	intelligence	=	knowledge	<-thinking-->	knowledge

Example for all three measures:

information	$I_{O1} = \{ (x_1, y_1), (\cancel{x_1}, y_2), (\cancel{x_2}, y_1), (x_2, y_2) \}$	=	$\text{Id } 4/2 = 1 \text{ Bit,}$
	$I_{O2} = \{ (\cancel{x_1}, y_1), (x_1, y_2), (x_2, y_1), (\cancel{x_2}, y_2) \}$	..	= 1 Bit,
	$I_{S1} = \{ (r_1, s_1), (\cancel{f_1}, s_2), (\cancel{f_2}, s_1), (r_2, s_2) \}$	..	= 1 Bit,
	$I_{S2} = \{ (\cancel{f_1}, s_1), (r_1, s_2), (r_2, s_1), (\cancel{f_2}, s_2) \}$	..	= 1 Bit,

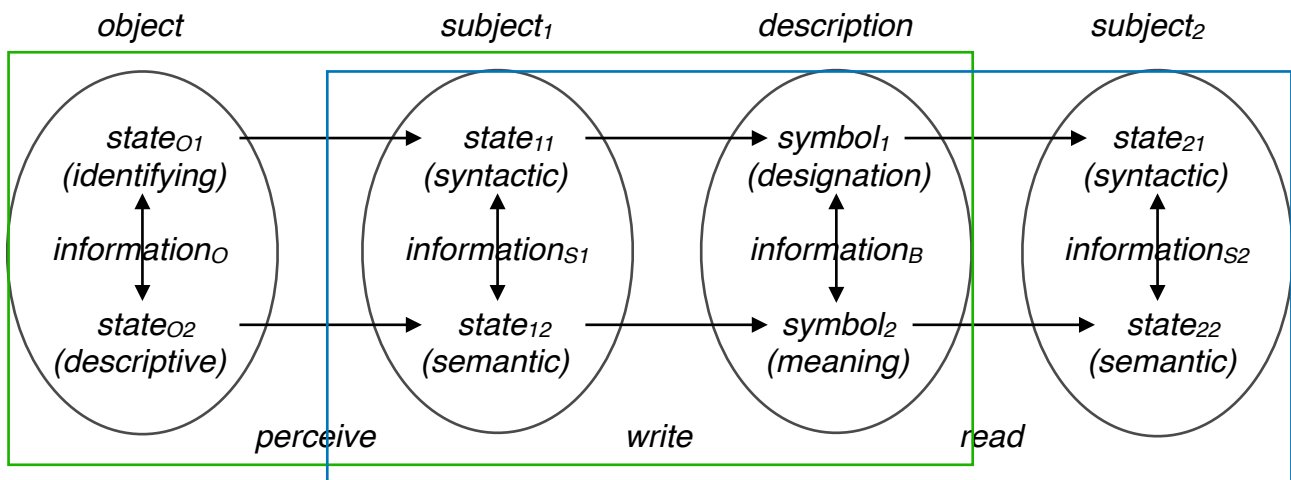
$$\begin{aligned}
\text{knowledge } I_{OS1} &= \{ (I_{O1}, I_{S1}), (I_{O1}, I_{S2}), (I_{O2}, I_{S1}), (I_{O2}, I_{S2}) \} & \dots & = 1 \text{ Bit,} \\
I_{OS2} &= \{ (I_{O1}, I_{S1}), (I_{O1}, I_{S2}), (I_{O2}, I_{S1}), (I_{O2}, I_{S2}) \} & \dots & = 1 \text{ Bit,} \\
\text{intelligence } I_{OS2} &= \{ (I_{OS1}, I_{OS1}), (I_{OS1}, I_{OS2}), (I_{OS2}, I_{OS1}), (I_{OS2}, I_{OS2}) \} & \dots & = 1 \text{ Bit.}
\end{aligned}$$

The search for unified science and world formula, as for intelligent science, is therefore to reduce the  $N_{XYRS12..M}$  in  $I_{XYRSM}$ .

## Describing

A third, important area of space next to objects and subjects are *descriptions*.

They take an intermediate position. On the one hand, they replace **objects** that the subject cannot reach spatially or temporally, are perceived when reading, and are actively generated when writing. On the other hand, they depict objects and thus, like **subjects**, store knowledge.



Descriptions are mostly sequences of symbols that are assigned to one another via their designations /16/.

The *syntax*, *semantics* and *pragmatics* of descriptions are intersubjectively agreed definitions.

The syntax embodies the structure of the symbol sequences, is information. The semantics of the symbol sequences, the meaning, is the knowledge. And the Pragmatics, i.e. efficiency, is intelligence.

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

The description of the description thus fulfills the recursive definition required above. The present text embodies intelligent knowledge as a hierarchy, through self-reference meta-knowledge /17/.

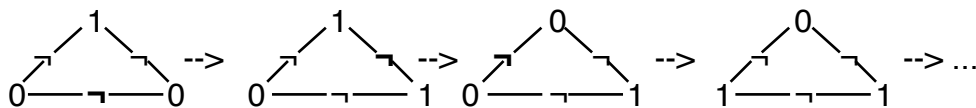


# Remarks

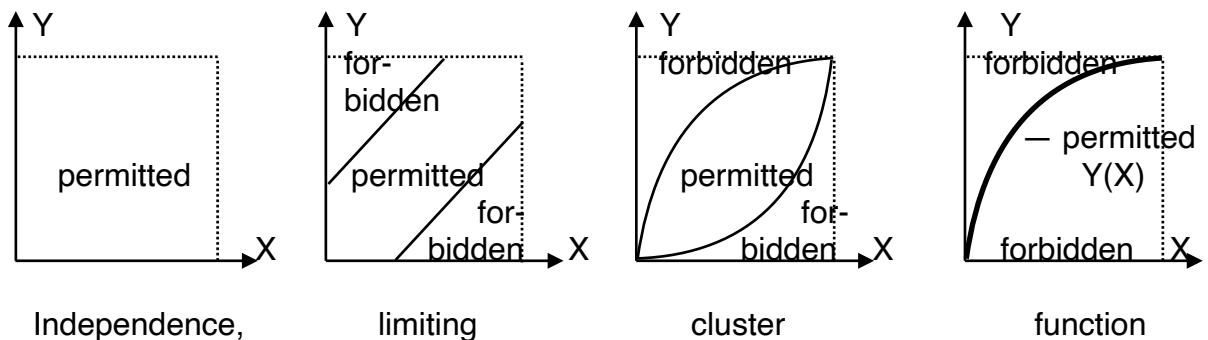
- 0:** The remarks have been moved out of the text for smoother reading design. Links in italics make it easier to jump back and forth. Other passages in italics refer to references or external sources, such as Wikipedia entries or specialist articles on the Internet.
- 1:** Value here is not in the sense of "imparting values", but as the content of the space/time elements or as a quantity of qualities. The concept of substance is too material, since abstract or subjective quantities are also involved. Examples of values are the mass "10 grams", the color "red", the price "3 euros" or the length "2 meters".
- 2:** Places, times, values, i.e. *entities*, are *relata* of the relation "different from". All the others, such as set, direction or dependency, build on this basic relation.
- 3:** Effect occurs as directed influence, like causality and function, or as undirected interactions such as correlation and structure.
- 4:** Dependent values change to avoid forbidden combinations. At contradiction between two values, however, is indeterminate which of the two changes. Then the majority dominates, like the example of negating dependency between the values 0 and 1 shows:

$$\begin{array}{rcl}
 0 \leftrightarrow 0 & \rightarrow & 0 \leftrightarrow 1 \text{ oder } 1 \leftrightarrow 0 \quad ? \\
 0 \leftrightarrow 0 \leftrightarrow 0 & \rightarrow & 0 \leftrightarrow 1 \leftrightarrow 0 \quad ! \\
 0 \leftrightarrow 0 \leftrightarrow 0 \leftrightarrow 0 \leftrightarrow 0 & \rightarrow & 0 \leftrightarrow 1 \leftrightarrow 0 \leftrightarrow 1 \leftrightarrow 0 \quad !
 \end{array}$$

From three contradictory values, this can lead to cyclical changes:



- 5:** The following state spaces show how the proportion of allowed combinations decreases and the forbidden increases:



**6:** Example of the *cross product* of sets (dimensions, properties):

$$\{ \text{good, average, bad} \} \times \{ \text{free, cheap, expensive} \} \times \{ \text{ring, car, house} \} = \{ (\text{good, free, ring}), (\text{good, free, car}), \dots, (\text{bad, expensive, house}) \},$$

as a tuple set, and to the *dyadic product* of tuples (objects, states):

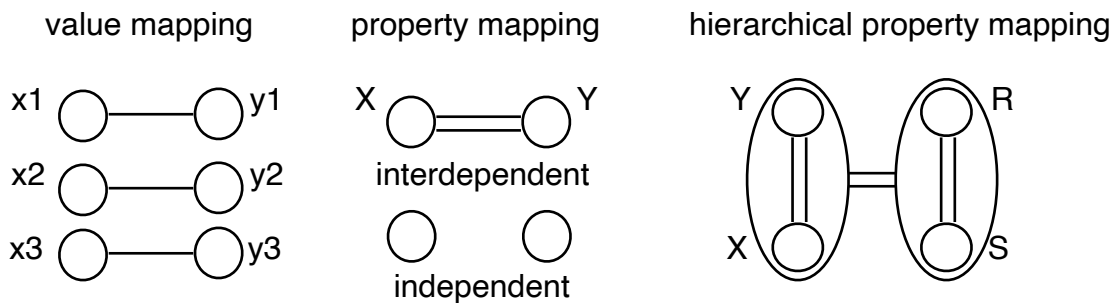
$$(\text{good, free, ring}) \otimes (\text{medium, cheap, car}) \otimes (\text{bad, expensive, house}) = ((\text{good, free, ring}), (\text{good, free, car}), \dots, (\text{bad, expensive, house})),$$

as tuple-tuple. But both form a table with three columns or three rows, which can be completed and limited.

**7:** If the flow of information directed by causality is to be measured, the states of X must be generated actively acting, to determine the resulting states of Y. Based on the spelling of Pearl (*Pearl 2009*), with "do(x)", this is notated as follows:

$$I_{\text{kausal}} := \text{Id} ( N_X * N_Y / N_{\text{do}(X)Y} ).$$

**8:** The representation of hierarchical relations becomes clearer if all values to one property and the strokes of all values to a double stroke be summarized. If the properties are independent and each value is connected to each other, there is no double stroke between properties.



**9:**  $I_{XYZ} = \text{Id} ( N_X * N_Y * N_Z / N_{XYZ} ) = \text{Id} ( ( N_X * N_Y * N_Z / N_{XYZ} ) * ( N_{XY} / N_{XY} ) )$   
 $= \text{Id} ( N_X * N_Y / N_{XY} ) + \text{Id} ( N_{XY} * N_Z / N_{XYZ} ) = I_{XY} + I_{(XY)Z}.$

**10:** If X and Y are independent, then  $N_{XY} = N_X * N_Y$  and  $I_{XY} = \text{Id} ( N_X * N_Y / N_X * N_Y ) = \text{Id} 1 = 0.$

But if X and Y are fully interdependent, then  $N_{XY} = N_X = N_Y = N$  and  $I_{XY} = \text{Id} ( N * N / N ) = I_{\text{max}} = \text{Id} N.$

All other cases are in between with  $N_{XY} \leq N_X * N_Y$  and  $0 \leq I_{XY} \leq I_{\text{max}}.$

**11:** *Smolin* questions the usefulness of information as a measure of dependency: "Information" is present whenever the values of two physical variables are correlated with each other. But we already have measures for correlations." But information as a measure of interdependence is more appropriate than others. So it also applies to non-linear, non-Gaussian, continuous or symbolic connections. *Terry Speed* calls information "A Correlation for the 21st Centuries".

**12:** The *mutual information* of Shannon's Communication Theory corresponds to our defined information measure. Our maximum information is his entropy  $H_x$

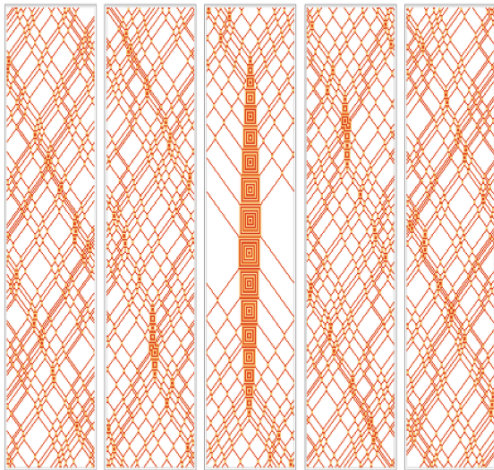
$$I_{\max} = I_{XX} = \text{Id } N_x = - \sum_x^{N_x} p_x \text{Id } p_x = H_x,$$

or his information of selecting one letter from an alphabet:

$$I_{\max} = I_{X1} = \text{Id } N_x * 1/1 = \text{Id } N_x = H_x.$$

**13:** Stephen *Wolfram* denies the existence of objective chance, even in quantum physics. He sees chance as observer-induced:

"Really all that matters is that the observer is computationally bounded. And it's then the basic computational mismatch between the observer and the computational irreducibility of the underlying system that leads us to "experience" the Second Law."



"A 'particle cellular automaton' like the one from the previous section has transformations that "go both ways", making its rule perfectly reversible."



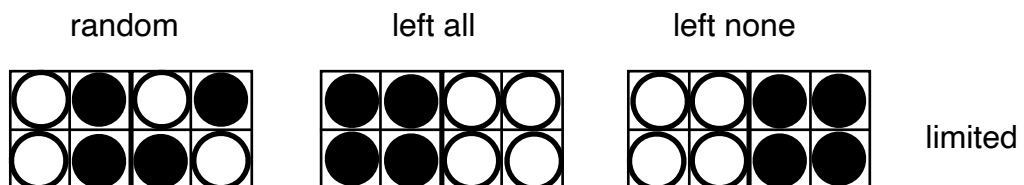
"To me it's a surprising and deeply beautiful scientific unification: that all three of the great foundational theories of physics—General Relativity, quantum mechanics and statistical mechanics—are in effect manifestations of the same core phenomenon: an interplay between computational irreducibility and our nature as observers."

**14:** Variable relations adapt to constant ones. The neurons of brains realize the relations - the values - and the variably connecting synapses the relations - the effects. The connection via synapses is directional. That may be one reason why often directed causality is assumed where only undirected correlation prevails

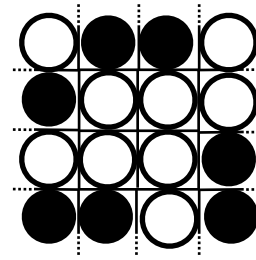
**15:** The probability of a two-tuple exceeds that of the three-tuple ones:

$$p(1) * p_1(2) \geq p(1) * p_1(2) * p_{12}(3) \text{ weil } 1 \geq p_{12}(3).$$

At the same time, the larger the area in which a value is sought, the larger is the probability (relative frequency) of finding it:



Frequency that left or center all (none):  
 black, white:  $p_S = p_W = 1/2$   
 $p_{LS} = p_{LW} = 4/8 * 3/7 * 2/6 * 1/5 = 1/70$   
 $p_{MS} = p_{MW} = 1/2 * 1/2 * 1/2 * 1/2 = 1/16 = 1/2^4$   
 black, white, red:  $p_S = p_W = p_R = 1/3$   
 $p_{LS} = p_{LW} = p_{LR} = 4/12 * 3/11 * 2/10 * 1/9 = 1/495$   
 $p_{MS} = p_{MW} = p_{MR} = 1/3 * 1/3 * 1/3 * 1/3 = 1/3^4$   
 N values with  $p_i = 1/N$   
 and M locations  $p_{Mi} = 1/N^M$   
 Probability that certain value i at least once in  
 region with M fields:  $p_{1i} = 1 - (1 - 1/N)^M$ .



Space  
unlimited

- 16:** The minimal, universal and *complete* programming language "Leibniz" served the search for the essence of languages and as an introduction to programming. The basis of the language are the symbols, for the values, and their recursive assignment, for the effects (*Sösemann 2010*).
- 17:** The idea of self-application was already suggested by the author (*Sösemann 1989*) by his "Programming of Programming".

## References

- *Pearl, Judea: Causal inference in statistics: An overview. University of California 2009.*
- *Shannon, Claude; Weaver, Warren: The Mathematical Theory of Communication. Urbana 1949.*
- *Smolin, Lee: Quantenwelt - Wie wir zu Ende denken, was mit Einstein begonnen hat. DVA 2019.*
- *Sösemann, Friedrich: Programmieren des Programmierens. 1989.*
- *Sösemann, Friedrich: Programmieren lernen mit "Leibniz". 2010.*
- *Speed, Terry: A Correlation for the 21st Century. in Science 2011.*
- *Wolfram, Stephen: "Computational Foundations for the Second Law of Thermodynamic". Stephen Wolfram Writings 2023.*

# Information, Wissen und Intelligenz als Relationen-Hierarchie

*Friedrich Sösemann 6/2023*

## Begriffe

*Welt, Relation, Information, Komplexität, Wahrnehmen, Lernen, Wissen, Denken, Intelligenz, Beschreiben.* ( [Link zur Anmerkung /0/](#) )

## Zusammenfassung

Information, Wissen und Intelligenz werden als Hierarchie von Relationen definiert: Information als abhängige Eigenschaften, Wissen als abhängige Informationen und Intelligenz als abhängiges Wissen. Für alle drei gilt das gleiche Abhängigkeits-Maß. Syntax, Semantik und Pragmatik von Beschreibungen verkörpern Information, Wissen und Intelligenz. Die Präzision und Messbarkeit dieser Begriffe soll Unschärfen und Widersprüche bei ihrer Anwendung verringern.

## Ansatz

*Information* ist in aller Munde. Obwohl schon vor über siebenzig Jahren von *Shannon* als beseitigte Unbestimmtheit formuliert, wird sie noch immer sehr unterschiedlich interpretiert. Das erschwert den Überblick und gegenseitigen Austausch. Für *Wissen* und *Intelligenz* fehlen einheitliche Definitionen und Maße sogar gänzlich.

Um dieses Problem zu lösen, wird hier eine präzise und allgemeine Definition dieser Begriffe angestrebt. Präzision und Allgemeinheit sollen durch Anwendung abstrakter mathematischer Mittel erreicht werden.

Doch dieser Lösungsansatz führt zu neuen Problemen:

Die Herleitung dieser Begriffe als Hierarchie von Relationen erfordert die Anwendung von Beschreibung, Wahrheit und Angemessenheit. Deren Anwendung schon vor ihrer Definition ist ein Teufelskreis.

Den Ausweg bieten rekursive Definitionen: Aus dem einfachsten Begriff werden jene abgeleitet, die dieser selbst voraussetzt. Damit schliesst sich der Kreis.

Der Preis der Allgemeinheit durch Abstraktion ist der Verlust von Details. Somit darf nur von unwesentlichen Merkmalen abstrahiert werden. Beispielsweise werden Information und Wissen als gerichtet empfunden, ihr Maß ist jedoch nicht gerichtet.

Exakte und universelle mathematische Begriffe reichen allein nicht aus. Ihnen fehlt die Bedeutung, ihre Zuordnung zur Welt. Daher muss die Welt zunächst durch eine möglichst minimale und universelle *Ontologie* definiert werden.

## Welt

Die *Welt* besteht aus *Werten* /1/ in Raum und Zeit und die Werte *wirken* aufeinander ein.

Der *Raum* ist das geordnete Nebeneinander von Orten, die *Zeit* das geordnete und gerichtete Nacheinander von Zeiten (Zeit-Punkten) /2/.

An einem Ort zu einer Zeit können mehrere Werte auftreten. Diese sind aber verschiedenen *Eigenschaften* (*Dimensionen*) zugeordnet. Die Werte einer Eigenschaft kommen niemals gemeinsam an einem Ort zu einer Zeit vor: Der Apfel ist rot oder grün, nie rot und grün, wohl aber oben rot und unten grün oder jetzt grün und später rot. Jeder Ort besitzt zu jeder Zeit je einen Wert aller Eigenschaften: Der Apfel ist rot, süß, rund und teuer.

Die Gesamtheit aller Werte eines Ortes oder von allen Orte eines Raum-Bereiches zu einer Zeit wird *Zustand* genannt.

Wenn die Anzahl von Orten, Bereichen oder Zeiten grösser als die unterschiedlicher Werte ist, treten die Werte mehrfach auf, besitzen verschiedene *Häufigkeiten*.

Die Wirkung /3/ zwischen direkt benachbarten Werten verhindert oder bevorzugt deren *Kombination*, beziehungsweise lassen diese seltener oder häufiger vorkommen. Durch die Ordnung von Raum und Zeit wird Nachbarschaft ermöglicht.

Nur Werte verschiedener Eigenschaften an einem Ort zu einer Zeit, wie rot und süß, oder benachbarte Werte gleicher Eigenschaft, wie rot neben grün, sind möglich.

*Abhängigkeit* erzeugt Veränderung /4/, aber auch *Strukturen*, *Gesetze* und *Regeln* /5/.

## Relation

Die mathematische *Relation* beschreibt die Abhängigkeit zweier Werte-Mengen als Untermenge ihres Produktes:

$$X = \{ x_1, \dots, x_{N_x} \}, Y = \{ y_1, \dots, y_{N_y} \}, R(X,Y) \subseteq X \otimes Y. \quad /6/$$

Die Mengen bilden Eigenschaften, ihre Elemente deren Werte und das Produkt erzeugt alle möglichen Kombinationen. Die Untermenge enthält dann nur die, trotz Abhängigkeit, verbleibenden Kombinationen.

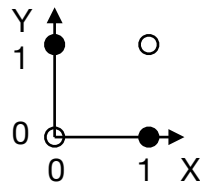
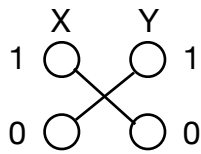
Bei Unabhängigkeit der Eigenschaften sind alle Kombination erlaubt, die *Mächtigkeit* der Relation ist dann das Produkt der Mächtigkeit der Ausgangsmengen:

$$|R| = |X| * |Y|.$$

Bei vollständiger Abhängigkeit treten hingegen nur eineindeutige Zuordnungen auf:

$$|R| = |X| = |Y|.$$

Relationen werden auch durch Graphen, Diagramme oder Tabellen dargestellt. Das Beispiel zeigt die eineindeutige Relation der Negation binärer Grössen  $X = \neg Y$ :



X	Y
0	1
1	0

## Information

Information sei die Abhängigkeit von Eigenschaften.

Die Abhängigkeit kann gerichtet, wie bei "Information über", oder nicht gerichtet, wie bei Strukturen, sein. Ihr Maß ist aber ungerichtet /7/ und lässt so alle Möglichkeiten offen.

$$I_{XY} = \text{Id} ( N_X \cdot N_Y / N_{XY} ), \quad N_X \text{ --- } N_Y \quad /8/$$

sei das Maß der Abhängigkeit zwischen den Eigenschaften X und Y, mit  $N_X \cdot N_Y$  als Mächtigkeit der Produktmenge  $X \otimes Y$ , sowie  $N_{XY}$  als Mächtigkeit der Untermenge  $R(X,Y)$ . Deren Quotient beschreibt die Abhängigkeit: Steigt diese, so verringert sich die Zahl tatsächlicher Kombinationen im Nenner, der Betrag des Quotienten wächst.

Damit die Information bei Unabhängigkeit Null ist und weitere Eigenschaften addiert werden können /9/, obwohl sich die Anzahlen multiplizieren, wird der Logarithmus des Quotienten verwendet /10/. Der binäre Logarithmus "ld" führt zur Maßeinheit "Bit".

Beim Beispiel der Negation binärer Eigenschaften werden die vier mögliche Wert-Zuordnungen auf zwei eingeschränkt. Das ergibt  $I_{XY} = \text{ld} 2 \cdot 2 / 2 = \text{ld} 2 = 1$  Bit.

Damit unterscheidet sich die Information nur durch ihr Maß von anderen Korrelations-Maßen /11/.

Die Information aller Werte einer Eigenschaft, die maximale oder Selbstinformation, ergibt mit  $N_X = N_Y = N_{XY} = N$ :

$$I_X = \text{ld} ( N \cdot N / N ) = \text{ld} N. \quad /12/$$

Abhängigkeit kann nicht nur die Anzahl zugeordneter Werte verringern, sondern auch nur deren Häufigkeit verändern. Dann wird statt der Gesamtzahl N die Häufigkeit der Werte  $p_n$  verwendet. Die Häufigkeit des tatsächlichen gemeinsamen Auftretens  $p_{xy}$  wird zu der bei Unabhängigkeit  $p_x \cdot p_y$  ins Verhältnis gesetzt. Das Informationsmaß, zweier ungleich verteilter Eigenschaften, ist dann der gewichtete Mittelwert über alle Kombinationen:

$$I_{XY} = \sum_x \sum_y p_{xy} \cdot \text{ld} ( p_{xy} / p_x \cdot p_y ).$$

Information ist auch für kontinuierliche Eigenschaften definiert. Die diskreten Häufigkeiten  $p_{xy}$  werden durch kontinuierliche Dichten  $P(x,y)$  und die Summen durch Integrale ersetzt:

$$I_{XY} = \int \int P(x,y) * \text{Id} ( P(x,y) / P(x) P(y) ) dx dy.$$

## Komplexität

Die Information von mehr als zwei Eigenschaften eines Bereiches wird auch *Komplexität* genannt. Bei ihr ist die Vielfalt vorgeschlagener Maße noch größer als bei der Information.

Dabei erfordert ihre Definition lediglich die Erweiterung von zwei auf M Größen (analog zu /9/):

$$I_M = \text{Id} ( N_1 * N_2 * .. * N_M / N_{12..M} ).$$

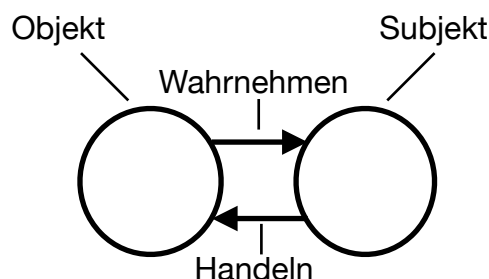
Die Information innerhalb eines Bereiches wird "intern", außerhalb "extern", oder auch "aktuell" und "potentiell" genannt. Beider Summe ist die maximale Information:

$$\begin{aligned} I_{\max} &= I_{\text{int}} + I_{\text{ext}} = I_{\text{akt}} + I_{\text{pot}} \\ &= \text{Id} ( N_1 * N_2 * .. * N_M / N_{12..M} ) + \text{Id} N_{12..M} \\ &= \text{Id} N_1 * N_2 * .. * N_M. \end{aligned}$$

Beim Beispiel der Negation gilt:  $I_{\max} = \text{Id} 2^2 = 2$  Bit,  $I_{\text{int}} = \text{Id} 2^2/2 = 1$  Bit,  $I_{\text{ext}} = \text{Id} 2 = 1$  Bit.

## Wahrnehmen

Wahrnehmung ist die gerichtete Abhängigkeit zwischen *Objekt* und *Subjekt* oder zwischen einem Subjekt und seiner Umgebung. *Wahrnehmen* und *Handeln* erzeugen die Abhängigkeit beider:



Wahrgenommen werden Werte. Ihr Zusammenfassen zu Eigenschaften erfolgt im Subjekt und wird durch unterschiedliche Sinne, wie Sehen, Hören, Fühlen, grob vorgegeben.

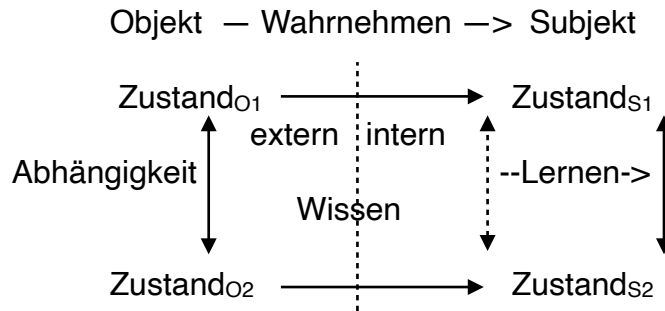
Wenn sich Handeln und Wahrnehmen abwechseln, wird gezieltes Lernen ermöglicht, aber auch das externe Herstellen gewünschter Objekte, sogenannter *Artefakte*.

Subjekte nehmen nur einen Teil der Werte ihrer Umgebung wahr. Ihre Komplexität ist notwendig geringer als die der Welt. Das bewirkt dann den Zufall aus Unwissenheit /13/.



# Lernen

Lernen schafft Wissen: Werden nicht nur einzelne Werte, sondern mehrere und deren Vorhandensein oder Fehlen wahrgenommen, so werden die Zustände des Objektes erfasst und auf Zustände des Subjektes, die Empfindungen, abgebildet:



Räumliche oder zeitliche Bewegung des Subjektes ermöglicht es, verschiedene Werte wahr zu nehmen. Lernen ersetzt dann die transitive Korrelation durch Kausalität und speichert damit externe Informationen intern ab.

*Speichern* verwandelt Nacheinander in Nebeneinander, transformiert Zeit in Raum.

Objekte und Subjekte unterscheiden sich durch die Variabilität ihrer Relationen. Diese ist extern konstant, intern aber zunächst variabel und durch Lernen konstant. Die Relationen der Subjekte passen sich denen der Objekte an /14/.

Das Beispiel zeigt, wie die feste Negations-Relation des Objektes über die Identitäten der Wahrnehmung im zunächst unabhängigen und variablen Subjekt realisiert wird:

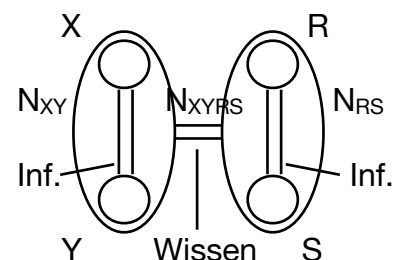
Objekt		Wahrnehmung		Subjekt			
X	Y	X	R	Y	S	R	S
0	1	0	0	0	0	<del>0</del>	<del>0</del>
1	0	1	1	1	1	0	1
						1	0
						<del>1</del>	<del>1</del>

# Wissen

Wissen sei die Abhängigkeit von Informationen.

Wahrgenommenes und handelndes Wissen sind gerichtet. Sein Maß, die Abhängigkeit zwischen Objekt und Subjekt, jedoch nicht:

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



mit  $N_{XY}$  als Anzahl der externen und  $N_{RS}$  der internen Informationen;  $N_{XYRS}$  ist die Anzahl der tatsächlichen Zuordnungen beider.

Die Summe der Komplexität von Objekt, Subjekt und Wissen ergibt die Gesamtkomplexität

$$I_{XY} + I_{RS} + I_{XYRS} = \text{Id } N_X * N_Y / N_{XY} * N_R * N_S / N_{RS} * N_{XY} * N_{RS} / N_{XYRS}$$

$$= \text{Id } N_X * N_Y * N_R * N_S / N_{XYRS} = I_{\text{ges.}}$$

## Denken

Das Denken verknüpft Wissen und schafft damit Intelligenz.

Denken bildet Eigenschaften aus Werten, Gesetze und Regeln aus Zuständen, Strukturen aus Elementen und *Schlüsse* aus Annahmen.

Beispiele dazu sind: Rot und Grün sind Farben; 01 und 10, nicht aber 00 und 11, definieren die Negation; Äpfel sind grün und sauer oder rot und süß;  $a < b$  und  $b < c$  gibt  $a < c$ .

Wahrgenommenes Wissen ist elementar, besteht aus getrennten Fakten, aus der Verbindung externer und interner Informationen.

Einfache Fakten sind im externen physikalischen Raum schneller zu finden als zusammengesetzte /15/. Somit werden bevorzugt Fakten wahrgenommen und gelernt. Erst im Subjekt werden sie durch Denken zu Theorien vereint.

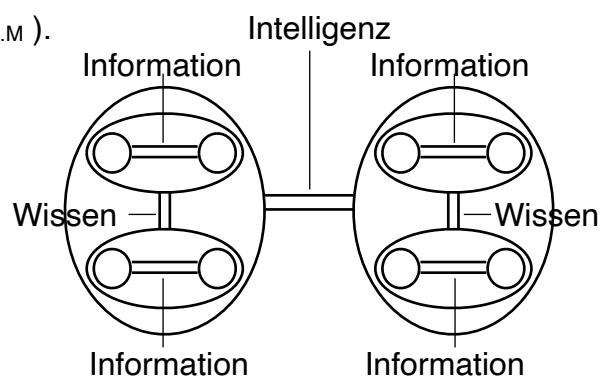
Im physikalischen Raum können nur Werte am gleichen Ort zur gleichen Zeit oder direkt benachbarte aufeinander einwirken (*Lokalität*). Im Subjekt sind hingegen alle Werte benachbart, alle Neuronen können über *Axone* miteinander verbunden werden, der *Merkmalsraum* ersetzt den physikalischen Raum.

## Intelligenz

*Intelligenz* sei die Abhängigkeit von Wissen, deren Komplexität.

Sie und ihr Maß sind ungerichtet. Ihr Maß, ist die *Komplexität* des Wissens, die Reduktion der Zahl der Verbindungsmöglichkeiten aller Wissensselemente:

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



Ebene 1:	Information	=	Zustand	<—Wirken—>	Zustand
Ebene 2:	Wissen	=	Information	<—Wahrnehmen—>	Information
Ebene 3:	Intelligenz	=	Wissen	<—Denken—>	Wissen

Beispiel für alle drei Maße:

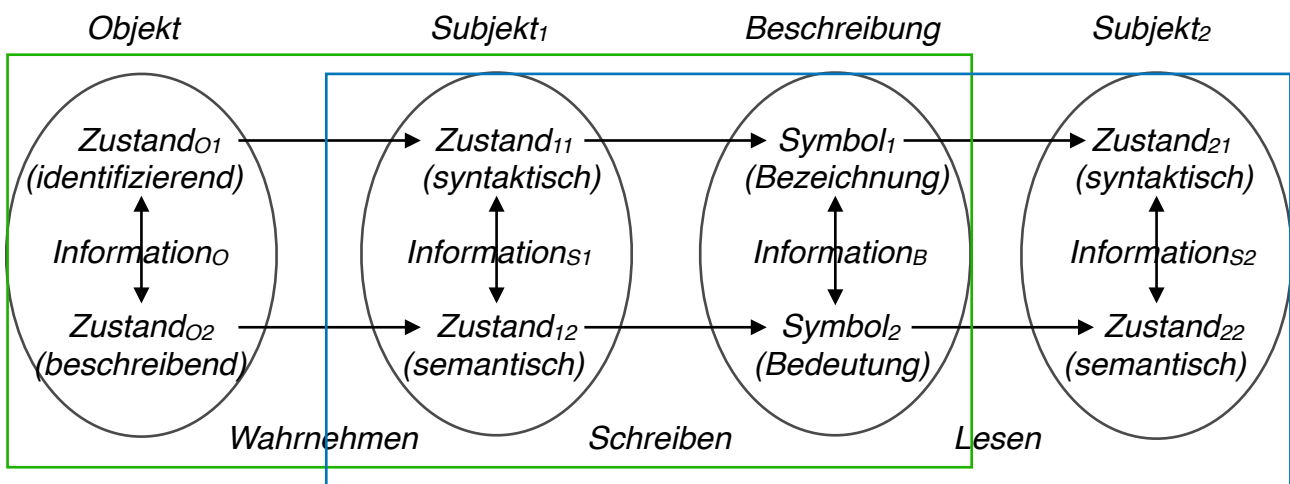
Information	$l_{O1} = \{ (x_1, y_1), (\cancel{x_1}, y_2), (\cancel{x_2}, y_1), (x_2, y_2) \}$	= $\log_2 4 = 2$ = 1 Bit,
	$l_{O2} = \{ (\cancel{x_1}, y_1), (x_1, y_2), (x_2, y_1), (\cancel{x_2}, y_2) \}$	.. = 1 Bit,
	$l_{S1} = \{ (r_1, s_1), (\cancel{r_1}, s_2), (\cancel{r_2}, s_1), (r_2, s_2) \}$	.. = 1 Bit,
	$l_{S2} = \{ (\cancel{r_1}, s_1), (r_1, s_2), (r_2, s_1), (\cancel{r_2}, s_2) \}$	.. = 1 Bit,
Wissen	$l_{OS1} = \{ (l_{O1}, l_{S1}), (\cancel{l_{O1}}, l_{S2}), (\cancel{l_{O2}}, l_{S1}), (l_{O2}, l_{S2}) \}$	.. = 1 Bit,
	$l_{OS2} = \{ (\cancel{l_{O1}}, l_{S1}), (l_{O1}, l_{S2}), (l_{O2}, l_{S1}), (\cancel{l_{O2}}, l_{S2}) \}$	.. = 1 Bit,
Intelligenz	$l_{OS2} = \{ (l_{OS1}, l_{OS1}), (\cancel{l_{OS1}}, l_{OS2}), (\cancel{l_{OS2}}, l_{OS1}), (l_{OS2}, l_{OS2}) \}$	.. = 1 Bit.

Die Suche nach der Einheitswissenschaft und Weltformel, also intelligenter Wissenschaft, will das  $N_{XYRS12..M}$  in  $l_{XYRSM}$  reduzieren.

## Beschreiben

Ein dritter, wichtiger Raum-Bereich neben Objekten und Subjekten, sind *Beschreibungen*.

Sie nehmen eine Zwischenstellung ein. Einerseits ersetzen sie für das Subjekt räumlich oder zeitlich unerreichbare Objekte, werden beim Lesen wahrgenommen, beim Schreiben handelnd erzeugt. Andererseits bilden sie Objekte ab und speichern damit, wie Subjekte, Wissen:



Beschreibungen sind zumeist Folgen von Symbolen, welche über ihre Bezeichnungen einander zugeordnet sind /16/.

Die *Syntax*, *Semantik* und *Pragmatik* von Beschreibungen sind intersubjektiv vereinbarte Definitionen.

Dabei verkörpert die Syntax - die Struktur - ihre Information, die Semantik - die Bedeutung - ihr Wissen und die Pragmatik - die Effizienz - ihre Intelligenz.

Das Beschreiben der Beschreibung erfüllt damit die vorn geforderte rekursive Definition. Der vorliegende Text verkörpert als Hierarchie intelligentes Wissen, durch Selbstbezug Meta-Wissen /17/.

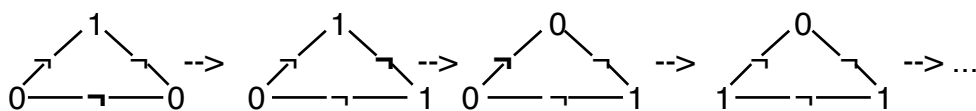
Ebene	Begriff	Maß	Beschreibung
0	Wert, Wirkung	$N_1, N_2, N_{12}$	Wörter
1	Information	$I_{XY} = \text{Id}(N_X * N_Y / N_{YX})$	Syntax
2	Wissen	$I_{XYRS} = \text{Id}(N_{XY} * N_{RS} / N_{XYRS})$	Semantik
3	Intelligenz	$I_M = \text{Id}(N_1 * N_2 * .. * N_M / N_{12..M})$	Pragmatik

# Anmerkungen

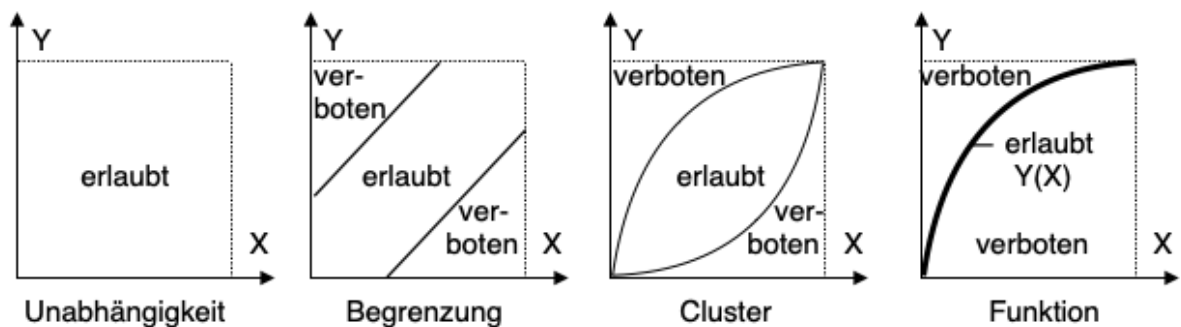
- 0:** Die Anmerkungen sind aus dem Text verlagert, um das Lesen dort flüssiger zu gestalten. Dabei erleichtern kursive Links die Hin- und Rücksprünge. Andere kursive Textstellen verweisen auf Literaturstellen oder externen Quellen, wie Wikipedia-Einträge oder Fachartikeln im Internet.
- 1:** Wert ist hier nicht im Sinne von "Werte vermitteln" oder "Werteunion", sondern als der Inhalt der Raum-/Zeit-Elemente oder als Quantität einer Qualität gemeint. Der Begriff *Substanz* ist zu materiell, da auch abstrakte oder subjektive Quantitäten umfasst werden sollen. Beispiele von Werten sind die Masse "10 Gramm", die Farbe "rot", der Preis "3 Euro" oder die Länge "2 Meter".
- 2:** Orte, Zeiten, Werte, also *Entitäten*, sind Relata der Relation "verschieden von". Auf dieser grundlegenden Relation bauen alle anderen, wie Menge, Richtung oder Abhängigkeit, auf.
- 3:** Die Wirkung tritt als gerichteter Einfluss, wie Kausalität und Funktion, oder als ungerichtete Wechselwirkung, wie Korrelation und Struktur, auf.
- 4:** Abhängige Werte ändern sich um verbotene Kombinationen zu vermeiden. Bei Widerspruch zwischen zwei Werten ist jedoch unbestimmt, welcher der beiden wechselt. Dann dominiert die Mehrheit, wie das Beispiel negierender Abhängigkeit zwischen den Werten 0 und 1 zeigt:

$$\begin{array}{rcl}
 0 \text{ --} \neg \text{--} 0 & \text{-->} & 0 \text{ --} \neg \text{--} 1 \text{ oder } 1 \text{ --} \neg \text{--} 0 \quad ? \\
 0 \text{ --} \neg \text{--} 0 \text{ --} \neg \text{--} 0 & \text{-->} & 0 \text{ --} \neg \text{--} 1 \text{ --} \neg \text{--} 0 \quad ! \\
 0 \text{ --} \neg \text{--} 0 \text{ --} \neg \text{--} 0 \text{ --} \neg \text{--} 0 \text{ --} \neg \text{--} 0 & \text{-->} & 0 \text{ --} \neg \text{--} 1 \text{ --} \neg \text{--} 0 \text{ --} \neg \text{--} 1 \text{ --} \neg \text{--} 0 \quad !
 \end{array}$$

Ab drei widersprüchlichen Werten kann das zu zyklischer Veränderung führen:



- 5:** Die folgenden Zustands-Räume zeigen, wie, beginnend mit Unabhängigkeit, über Begrenzung und Cluster bis Funktion, der Anteil erlaubter Kombinationen ab- und verbotener zunimmt.



**6:** Beispiel zum *Kreuzprodukt* von Mengen (Eigenschaften, Dimensionen):  
 $\{ \text{gut, mittel, schlecht} \} \times \{ \text{geschenkt, billig, teuer} \} \times \{ \text{Ring, Auto, Haus} \} =$   
 $\{ (\text{gut, geschenkt, Ring}), (\text{gut, geschenkt, Auto}), \dots, (\text{schlecht, teuer, Haus}) \},$

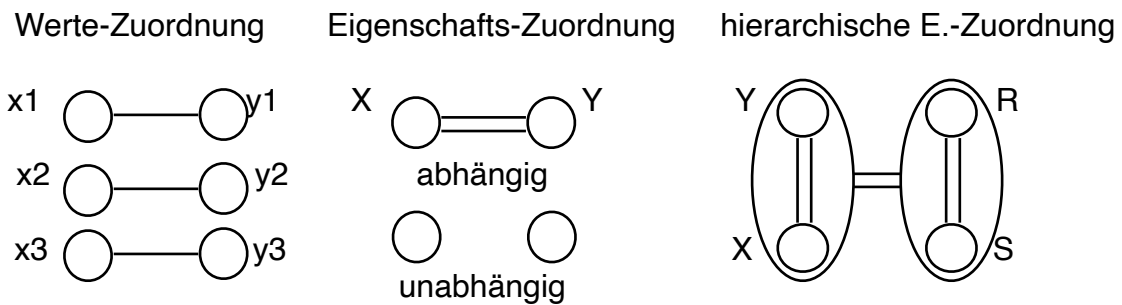
als Tupel-Menge, und zum *Dyadischen Produkt*: von Tupeln (Objekte, Zustände):  
 $(\text{gut, geschenkt, Ring}) \otimes (\text{mittel, billig, Auto}) \otimes (\text{schlecht, teuer, Haus}) =$   
 $((\text{gut, geschenkt, Ring}), (\text{gut, geschenkt, Auto}), \dots, (\text{schlecht, teuer, Haus})),$

als Tupel-Tupel. Doch beide bilden eine Tabelle mit drei Spalten oder drei Zeilen, die vervollständigt und begrenzt werden kann.

**7:** Wenn der durch Kausalität gerichtete Informationsfluss gemessen werden soll, müssen die Zustände von X aktiv handelnd erzeugt werden, um die sich daraus ergebenden Zustände von Y zu ermitteln. In Anlehnung an die Schreibweise von Pearl (*Pearl 2009*), mit „do(x)“, wird das wie folgt notiert:

$$I_{\text{kausal}} := \text{Id} ( N_X * N_Y / N_{\text{do}(X)Y} ).$$

**8:** Die Darstellung hierarchischer Relationen wird übersichtlicher, wenn alle Werte zur Eigenschaft und die Striche aller Werte- zu einem Doppelstrich der Eigenschafts-Zuordnung zusammengefasst werden. Bei Unabhängigkeit der Eigenschaften ist zwar jeder Wert mit jedem anderen verbunden, die Eigenschaften jedoch nicht.



**9:**  $I_{XYZ} = \text{Id} ( N_X * N_Y * N_Z / N_{XYZ} ) = \text{Id} ( ( N_X * N_Y * N_Z / N_{XYZ} ) * ( N_{XY} / N_{XY} ) )$   
 $= \text{Id} ( N_X * N_Y / N_{XY} ) + \text{Id} ( N_{XY} * N_Z / N_{XYZ} ) = I_{XY} + I_{(XY)Z}.$

**10:** Bei Unabhängigkeit von X und Y gilt mit  $N_{XY} = N_X * N_Y$

$$I_{XY} = \text{Id} ( N_X * N_Y / N_X * N_Y ) = \text{Id} 1 = 0.$$

Sind X und Y voll voneinander abhängig, ist mit  $N_{XY} = N_X = N_Y = N$

$$I_{XY} = \text{Id} ( N * N / N ) = I_{\text{max}} = \text{Id} N.$$

Alle anderen Fälle liegen mit  $N_{XY} \leq N_X * N_Y$  dazwischen

$$0 \leq I_{XY} \leq I_{\text{max}}.$$

**11:** *Smolin* zweifelt den Nutzen von Information als Maß der Abhängigkeit an: „»Information« ist jedes Mal vorhanden, wenn die Werte zweier physikalischer Variablen miteinander korreliert sind. Aber wir haben bereits Maße für Korrelationen.“

Aber Information als Maß gegenseitiger Abhängigkeit ist geeigneter als andere. So gilt sie auch für nichtlineare, nichtgaußsche, kontinuierliche oder symbolische Zusammenhänge. Terry Speed nennt Information „A Correlation for the 21st Century“.

- 12:** Dem definierten Informationsmaß entspricht die Transinformation der klassischen Kommunikationstheorie von Shannon. Seiner Entropie  $H_X$  entspricht die maximale Information:

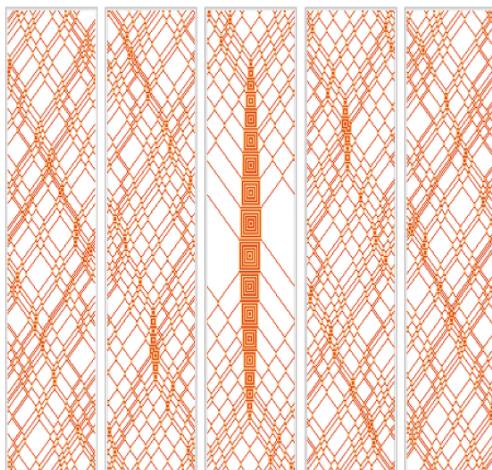
$$I_{\max} = I_{XX} = \text{Id } N_X = - \sum_x p_x \text{Id } p_x = H_X,$$

oder die Information der Auswahl eines Buchstabens aus einem Alphabet:

$$I_{\max} = I_{X1} = \text{Id } N_X^{*1/1} = \text{Id } N_X = H_X.$$

- 13:** Stephen Wolfram bestreitet die Existenz des objektiven Zufalls, sogar in der Quantenphysik. Er sieht Zufall als beobachterinduziert:

"Really all that matters is that the observer is computationally bounded. And it's then the basic computational mismatch between the observer and the computational irreducibility of the underlying system that leads us to "experience" the Second Law."



"A 'particle cellular automaton' like the one from the previous section has transformations that "go both ways", making its rule perfectly reversible."



"To me it's a surprising and deeply beautiful scientific unification: that all three of the great foundational theories of physics—General Relativity, quantum mechanics and statistical mechanics—are in effect manifestations of the same core phenomenon: an interplay between computational irreducibility and our nature as observers."

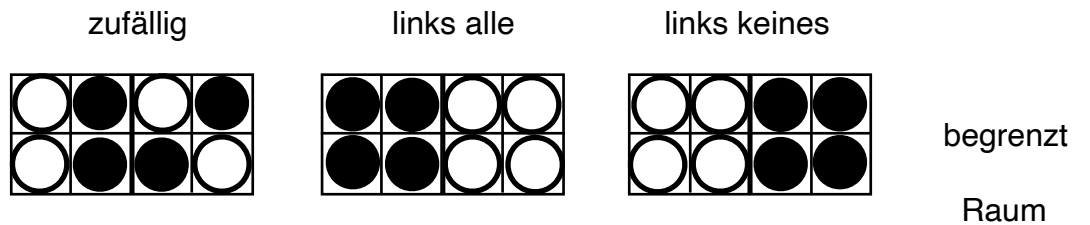
- 14:** Variable Relationen passen sich an konstante an. Die *Neuronen* von Gehirnen realisieren die Relata - die Werte - und die variabel verbindenden Synapsen die Relationen - die Wirkungen.

Die Verbindung über Synapsen ist gerichtet. Das mag ein Grund sein, warum oft gerichtete Kausalität vermutet wird, wo nur ungerichtete Korrelation herrscht.

- 15:** Die Wahrscheinlichkeit eines Zweier- übersteigt die des ergänzenden Dreiertupels:

$$p(1)*p_1(2) \geq p(1)*p_1(2)*p_{12}(3) \text{ weil } 1 \geq p_{12}(3).$$

Zugleich gilt: Je grösser der Bereich, in dem ein Wert gesucht wird, umso grösser ist die Wahrscheinlichkeit (relative Häufigkeit) ihn zu finden:



Häufigkeit, dass links oder mittig alle (keines):

schwarz, weiss:  $p_S = p_W = 1/2$

$p_{LS} = p_{LW} = 4/8 * 3/7 * 2/6 * 1/5 = 1/70$

$p_{MS} = p_{MW} = 1/2 * 1/2 * 1/2 * 1/2 = 1/16 = 1/2^4$

schwarz, weiss, rot:  $p_S = p_W = p_R = 1/3$

$p_{LS} = p_{LW} = p_{LR} = 4/12 * 3/11 * 2/10 * 1/9 = 1/495$

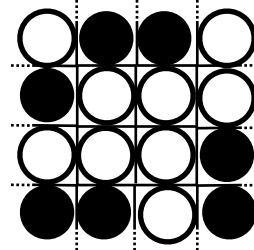
$p_{MS} = p_{MW} = p_{MR} = 1/3 * 1/3 * 1/3 * 1/3 = 1/3^4$

N Werte mit  $p_i = 1/N$

und M Orte  $p_{Mi} = 1/N^M$

Wahrscheinlichkeit, dass bestimmter Wert i wenigstens einmal in Bereich mit M

Feldern auftritt:  $p_{1i} = 1 - (1 - 1/N)^M$ .



unendlich

- 16:** Die minimale, universelle und *vollständige* Programmiersprache "Leibniz" (Sösemann 2010) diene der Suche nach dem Wesenskern von Sprachen und der Einführung in das Programmieren. Die Basis der Sprache bilden die Symbole, für die Werte, und deren rekursive Zuordnung, für die Wirkungen.
- 17:** Der Idee der Selbstanwendung wurde vom Autor (Sösemann 1989) schon beim "Programmieren des Programmierens" verfolgt.

## Literatur

- Pearl, Judea: Causal inference in statistics: An overview. University of California 2009.
- Shannon, Claude; Weaver, Warren: The Mathematical Theory of Communication. Urbana 1949.
- Smolin, Lee: Quantenwelt - Wie wir zu Ende denken, was mit Einstein begonnen hat. DVA 2019.
- Sösemann, Friedrich: *Programmieren des Programmierens*. 1989.
- Sösemann, Friedrich: *Programmieren lernen mit "Leibniz"*. 2010.
- Speed, Terry: A Correlation for the 21st Century. in Science 2011.
- Wolfram, Stephen: "Computational Foundations for the Second Law of Thermodynamic". *Stephen Wolfram Writings* 2023.