Anticipation, an Optimisation Principle for Cognition

Hanns Sommer
Department of Mechanical Engineering (FB 15), University of Kassel (GHK)
Mönchebergstr. 7, 34109 Kassel, Germany
Phone: 0561 8043261, Fax: 0561 8047768
email: sommer@rts.maschinenbau.uni-kassel.de

Abstract: Is cognition possible without a priori assumptions ?
Many philosophers and physicists claim that a reference frame for the ordering of the measurements is needed, before cognition can start. Our objective is to reduce the a priori assumptions and to deduce a reference frame directly from empirical data with an Anticipation Principle: Put the measurements in such an order that the credibility of your forecasts will be maximised.
The structure of space and time, obtained with this principle provides an explication for some phenomena of quantum mechanics.
Keywords : Anticipation, cognition, belief, foundationalism, relational reality.

1 Preconditions for cognition

Our understanding of measurements from our world depends on reference frames in which the observations will be put in order. These reference frames are obtained differently in the philosophical schools. The a priori knowledge of Kant, Hegel's logic and in the view of constructivists, the preknowledge accumulated in the history of mankind, are the prerequisites for our understanding. But these prerequisites are not obtained from the empirical data, they form part of our prejudices and may obscure our access to the world.

In this paper we demonstrate that the following three categories are sufficient for cognition:

- Measurements from the World,
- Belief (formalised with Dempster-Shafers Belief Calculus),
- The Principle (A): Put the empirical information in an order, such that anticipation will be optimally enabled.

With this basic assumptions we are not only able to order the events of our world into a predetermined reference frame but also to deduce this reference frame itself.

International Journal of Computing Anticipatory Systems, Volume 14, 2004
Edited by D. M. Dubois, CHAOS, Liège, Belgium, ISSN 1373-5411 ISBN 2-930396-00-8

The concept of space and time can be introduced without any further preassumption. The knowledge we obtain exclusively from measurements shows structures that are well- known from quantum mechanics. The EPR-phenomenon with the impossibility of a faster-than-light communication and the quantum Zeno effect can in this way be deduced from elementary principles. The reality formed by our knowledge is a "relational reality" where one observed fact is real only in relation to its interactions with the whole system. As Jeeva Anandan noticed, the idea of a relational reality removes the antromorphic concepts from the Copenhagen interpretation and provides therefore the objectivity of our cognition.

2 Dempster Shafer's Belief Calculus

2.1 The meaning of belief

Dempster Shafer's belief calculus is based on Principle (B): "If an event had often occurred in the past, then its appearance in the future is believable."
This principle does not provide a complete description of human belief, as human belief depends on intentions, desires, moods, emotions, memory restrictions and many other special conditions of human life ([3]). But the principle (B) plays also an important role for humans. It can be understood as the simplest possibility for a definition of belief, the essence that remains if the consequences of special realisations of a belief-function had been put into parenthesis ([4]).

To formalise principle (B), we use Fuzzy language, where statements have truth values in $[0,1]$ and the operators "not", "and" and "or" are defined here for $a, b \in [0,1]$ by:

$$\neg a := 1 - a; \quad a \wedge b := a \cdot b; \quad a \vee b := 1 - (1-a) \cdot (1-b).$$

For a set X, interpreted as the set of symptoms, we assume that the presence of subsets of symptoms $A_i \subseteq X, (i = 1, \ldots, n)$ justifies our belief in some consequences. The importance of A_i is characterised by a value $m(A_i) \in [0,1]$.
For a subset $B \subseteq X$ of observed symptoms, our belief in the consequence is then given by the equation:

$$\text{Bel}(B) := \bigvee_{\substack{i=1 \\ A_i \subseteq B}}^{n} m(A_i) \tag{1}$$

(As we are only interested in relative believe degrees and not in the absolute values, our arguments are not changed by rescaling and therefore equation (1) provides an equivalent formulation of Dempster-Shafers Belief Calculus.)

2.2 Calculation of the most believable answer to a well defined question

A well defined question consists:

- Of a question formulated in completely defined terms with reference to real measurement values.
- The specification of the precision in which the answer is needed.

Example: "Will the temperature in a plant exceed 150°C during the next hour?" An answer to this question may have the form.: "During the next hour the temperature will be small." This answer can be represented by the Fuzzy expression (Fuzzy set): "Not time great (respective to one hour) or temperature small."

The calculation of the believability of an answer to a well-defined question is done in the following steps:

(a) Fuzzy numbers are defined to specify the precision needed in different orders of magnitude of the measurements. The measurements can now be characterized by their membership degree to tuples of these Fuzzy numbers. We collect these tuples in the set U.

Example continued: Near temperatures of 150°C and time intervals of one hour, very fine Fuzzy numbers are used for the scaling of the measurements, whereas coarser Fuzzy numbers would be sufficient in other regions. Each element of U, for example: "Time near 30 minutes and temperature near 100°C" characterises a Fuzzy subset of measurements.

(b) Let the measurements be grouped together into a set M and let M_T denote the set of values produced by a Turing machine (model T). The membership degrees for the elements of U to the sets M and M_T can be calculated.

(c) A correspondence-degree for the membership functions of the elements $u \in U$ to M and to M_T is given by Fuzzy set theory. The believability of the behaviour of T is defined by this correspondence degree.

(d) To obtain the believability $\mathrm{Bel}(T)$ of the model, we have to put the believability degree of the behaviour of a model into relation to the Kolmogorov complexity of T.

(e) The believability of an answer is given by its correspondence to the elements of U that belong to highly believable models.

With this procedure we can find the most believable answer to a well-defined question in the finite set of possible answers whose complexity is bounded by a fixed bound.

The calculus of Fuzzy logic enables a formal realisation of the indicated steps:

(a) Fuzzy-numbers are defined by functions $\bar{\omega} : \mathbb{R} \mapsto [0,1]$, where $\bar{\omega}^{-1}(\alpha) := \{\omega \in \mathbb{R} \mid \bar{\omega}(\omega) \in [\alpha, 1]\}$ is connected for every $\alpha \in \{0,1\}$.
A finite set U of tupels of Fuzzy numbers is selected such that measurements from M can be classified with sufficient precision.
(b) The membership degree of an element $(u_1, \ldots, u_n) \in U$ to the measurements M is defined by:

$$A(u_1, \ldots u_n) := \bigvee_{(m_1, \ldots m_n) \in M} u_1(m_1) \wedge \cdots \wedge u_n(m_n)$$

The membership degree of an element $(u_1, \ldots, u_n) \in U$ to the set of values M_T produced by a Turing machine T is defined by the agreement in corresponding times. For time measurements m_1 we obtain:

$$deg((u_1, \ldots, u_n) \in M_T) := \bigwedge_{(m_1, \ldots, m_n) \in M} \neg u_1(m_1) \vee (u_2(m_2) \wedge \cdots \wedge m_n(u_n))$$

(c) The correspondence degree between M and M_T is a measure for the credibility of the behaviour of the model represented by a Turing machine T:

$$\text{Bel}(\text{behaviour} \quad T) := \bigvee_{u \in U} deg(u \in M_T) \wedge A(u)$$

(d) From the credibility of the behaviour of a model T and from its Kolmogorov complexity $L_K(T)$, we obtain the belivability of the model T ([6]):

$$\text{Bel}(T) := \text{Bel}(\text{behaviour} \quad T)^{-L_K(T)\vee} \tag{2}$$

where $LK(t)$ denotes the length of a shortest text that defines the Turing table of T, and $a^{-k\vee} = b$ iff $a = b \vee b \vee \cdots \vee b$ (k factors).
(e) For one most believable model and an answer defined by a Fuzzy set $(v_1, \ldots, v_n)$, our belief in this answer corresponds to the value:

$$\text{Bel}((v_1, \ldots, v_n)) := \text{Bel}(T) \wedge \bigwedge_{(u_1, \ldots, u_n) \in U} (deg(u_1, \ldots, u_n) \in M_T) \vee (\neg(u_1 \subseteq v_1) \vee \cdots \vee \neg(u_n \subseteq v_n))$$

for more than one very believable model, the different most believable answers must be aggregated with aggregation operators ([9],[7]).

3 The meaning of reality and the dependece of space and time on measurements

3.1 The emergence of space and time

Reality emerges from the evaluation of measurements by an observer with reference to his questions and intentions. Reality means therefore the idea, we obtain from the world through the world. The preconditions of our reality are:

(a) Measurements, where a measurement is a real value which is produced in the observers brain or is obtained from an instument. Neither the meaning of the instruments nor that of the values they provide is knowen prior to the formation of our reality.

(b) Basic adjectives which characterise our relation to the world. For a formation of our physical reality we don't need adjectives like beautiful or ugly but only the assessments small, intermediate and great for the different sorts of measurements we obtain from an experiment.

(c) Intentions, by which we will be constituted as living beings. This is the most difficult part for an understanding of our everyday-life-world, because there is no direct access available to human intentions. But for the formation of our physical reality, our intentions can be summarised with **Postulate (A):** "Our understanding of the world should optimally enable forecasts of future events." The role of the question in the calculation of the best confirmed answer is now played by Postulate (A).

In the rest of this section a formal definition of this prerequisites will be given and we will deduce the basic elements of our language and our imagination of time and space.

(a) The empirical information is provided by measurements $M := \{m_k = (m_{1,k}, m_{2,k}, \ldots, m_{n,k}) \mid k = 1, \ldots, K\}$ whose meaning is a priori unknown. Their signification emerges through our understanding.

(b) Let the set of basic adjectives: tiny, small, intermediate, great and enormous be represented by Fuzzy membership functions $\mu_i : \mathbb{R} \to [0,1), (i = 1, \ldots, 5)$. A general property of a measurement tuple $m = (m_1, \ldots, m_n)$ will be represented by an expression:

$$s(m) := \sum_{i=1,j=1}^{5,n} \alpha_{i,j} \cdot \mu_i(m_j) \quad \text{with} \quad \alpha_{i,j} \geq 0 \quad \text{and} \quad \sum_{i=1,j=1}^{5,n} \alpha_{i,j} = 1,$$

or more precisely, using the attribute "very" that is represented by exponents in the Fuzzy language:

$$s(m) := \frac{1}{q} \cdot \sum_{i=1,k_1=1,\ldots,k_n}^{5,K_1,\ldots,K_n} \alpha_{i,k_1,\ldots,k_n} \Pi_{j=1}^{n} \mu_i(m_j)^{k_j} \tag{3}$$

$$\text{with} \quad \alpha_{i,k_1,\ldots,k_n} \in \mathbb{N} \cup \{0\} \quad \text{and} \quad \sum_{i=1,k_1=1,\ldots,k_n=1}^{5,K_1,\ldots,K_n} \alpha_{i.k_1,\ldots,k_n} = q$$

The set S of all adjectives is defined by "or" connections of general adjectives. The Kolmogorov complexity of an adjective is (in analogy to section 2):

$$L_K(s) := \text{length of a shortest text that defines} \quad s$$

(c) To optimise our possibility to make forecasts, points in space and time are defined in such a way that the principles (C1) and (C2) hold:

(C1) The existence of each point and its distinction from other points will be maximally confirmed by the empirical data.

A point p is specified in our language by a subset of adjectives $S_p \subseteq S$.
The membership function $\mu(p,m) := \bigwedge_{s \in S_p} s(m)$ provides the membership degree of a measurement $m \in M$ to the point p. p is therefore characterised by the tupel $(S_p, \mu(p, \bullet))$, where $\mu(p, \bullet)$ denotes a Fuzzy subset of M.
Each measurement $m \in M$ defines a Fuzzy subset of the adjectives S by: $\bar{m}(s) := s(m)$.
Let $A(S_p \subseteq \bar{m})$ denote the Fuzzy truth of the statement $S_p \subseteq \bar{m}$ or the value $\sup_{s \in S} \neg S_p(s) \vee \bar{m}(s)$.
With the given notations, the confirmation of a point p will be defined:

$$\begin{aligned} \mathrm{Bel}(p) &:= (\text{confirmation of } p \text{ by measurements belonging to } p) \\ &\wedge \neg(\text{confirmation of } p \text{ by measurements not belonging to } p) \qquad (4) \\ &= (\bigvee_{m \in M} \mu(p,m) \wedge A(S_p \subseteq \bar{m})) \wedge \neg(\bigvee_{m \in M} (1-\mu(p,m)) \wedge A(S_p \subseteq \bar{m})) \end{aligned}$$

The collection $\mathbb{W}$ of all points is called space and the confirmation of all these points by the measurements M is given by:

$$\mathrm{Bel}(\mathbb{W}) := \bigvee_{p \in \mathbb{W}} \mathrm{Bel}(p)$$

The complexity of the space $\mathbb{W}$ is:

$$L_K(\mathbb{W}) := \text{length of a shortest text that defines all} \quad S_p \quad \text{for} \quad p \in \mathbb{W}$$

The most believable space and time structure $\mathbb{W}_{opt}$ is selected by the expression:

$$\mathbb{W}_{opt} := \arg \underbrace{\max}_{\text{possible worlds } \mathbb{W}} \mathrm{Bel}(\mathbb{W})^{-L_k(\mathbb{W})\vee} \qquad (5)$$

As the set of adjectives, whose complexity is bounded by a fixed bound is finite, it is possible to find the most believable space and time structure defined by these adjectives. On the other hand, as the use of very complex adjectives provides only complex descriptions, once we have found a very well confirmed world $\mathbb{W}_{opt}$, there is no hope to obtain a better confirmed world.

(C2) To enable forecasts, we equip $\mathbb{W}_{opt}$ with the topology that relates the nearness of two points p and $\hat{p}$ to the similarity of their descriptions:

$$\mathrm{sim}(p,\hat{p}) := (\underbrace{\bigvee}_{s\in S_p, m\in M} s(m) \wedge \mu(\hat{p},m)) \wedge (\underbrace{\bigvee}_{s\in S_{\hat{p}}, m\in M} s(m) \wedge \mu(p,m))$$

$(1 - \mathrm{sim})$ satisfies reflexivity, symmetry and triangularity inequality and is therefore a measure for distance.

3.2 Structures in the most believable world

The structures of $\mathbb{W}_{opt}$ allow further interpretations:

Here and Now is fixed by the best confirmed point $p \in \mathbb{W}_{opt}$.

A measure for the change $\mathrm{ch}(p,\hat{p})$ between a point p and one of its neighbours $\hat{p}$ is defined by the shortest text that is necessary for a description of the changes in S_P to get $S_{\hat{p}}$ and for the reverse case.

A change operator Z is a transformation of the definition S_P of one point p into the definition of another real or hypothetical point $p*$ represented by a set $S_{p*} \subseteq S$: $Z(S_p) = S_{p*}$. The complexity of Z is defined by: $L_K(Z) := \mathrm{ch}(p, p*)$.

A restriction of the world can be obtained by a restriction of our observation to several sorts of measurements $\tilde{M}\mid_{\{j_1,\dots,j_L\}} :=$
$\{(\tilde{m}_{j_1},\dots,\tilde{m}_{j_L}) \mid \exists (m_1,\dots,m_n) \in M : \tilde{m}_{j_1} = m_{j_1},\dots,\tilde{m}_{j_L} = m_{j_L}\}$
and $\{j_1,\dots,j_L\} \subset \{1,\dots,n\}$.
Each adjective $s \in S$ can be restricted to $\tilde{M}\mid_{\{j_1,\dots,j_L\}}$, using equation (3) with $\mu_i(m_j) = 1$ for $j \notin \{j_1,\dots,j_L\}$ and $\tilde{\alpha}_{i,\tilde{k}_{j_1},\dots,\tilde{k}_{j_L}} := \sum_{\tilde{k}_{j_1}=k_{j_1},\dots,\tilde{k}_{j_L}=k_{j_L}} \alpha_{i,k_1,\dots,k_n}$.
Let $\tilde{S}_P$ denote the restriction of S_p to $\tilde{M}\mid_{\{j_1,\dots,j_L\}}$.

Time is defined by the best confirmed uniform behaviour of a restricted world. This implies, the selection of a restriction of the set of measurements $\tilde{M}\mid_{\{j_1,\dots,j_L\}}$, a sequence of points $p_1,\dots,p_\kappa \in \mathbb{W}_{opt}$ and integers $l_1,\dots,l_\kappa \in \mathbb{N}$ such that the expression:

$$(\mathrm{sim}(Z^{l_1}\tilde{S}_p, \tilde{S}_{p_1}) \vee \mathrm{sim}(Z^{l_2}\tilde{S}_p, \tilde{S}_{p_2})\vee,\dots,\vee\mathrm{sim}(Z^{l_\kappa}\tilde{S}_p, \tilde{S}_{p_\kappa}))^{-L_K(Z)\vee}$$

will be maximised. The change operator Z^T that maximises this expression, defines the direction of time in $\mathbb{W}_{opt}$. $\text{ch}(\tilde{p}, \tilde{p}*) := \text{ch}(\tilde{S}_p, \tilde{S}_{p*})$ is a measure for the time that flew between p and $p*$.

The **dimension of the space** $\mathbb{W}_{opt}$ in a point p can be characterised by the value:

$$\log_3 \left(\frac{\| \{\tilde{p} \in \mathbb{W}_{opt} \mid \exists \breve{p} \in \mathbb{W}_{opt} : \quad \text{sim}(p, \breve{p}) > \varepsilon \quad \text{and} \quad \text{sim}(\breve{p}, \tilde{p}) > \varepsilon\} \|}{\| \{\tilde{p} \in \mathbb{W}_{opt} \mid \text{sim}(p, \tilde{p}) > \varepsilon\} \|} \right)$$

where $\| S \|$ denotes the number of elements of a set S and ε is chosen such that the dimension remains stable respective to small variations of ε.

$\mathbb{W}_{opt}$ is the **phase space** for the experiment which produces the measurements $m \in M$.

3.3 The most believable world forms a Relational Reality

(a) A comparison of different accesses to reality: In our view, understanding and reality are interwoven. Without understanding, no cognition is possible, which is free from prejudices that are not empirically justified.

The use of probability theory for a foundation of physical laws, needs first a description of the events, whose frequency should be detected. In this access to reality, the description language is prior to the discovery of probabilities and their dependencies. Physical laws are therefore dependent from a bootstrapping process, for which a first language will be needed. In the view of classical mechanics, an experiment takes place in a completely known environment. All instruments, their locations and the signification of the measurements have to be well-known for an interpretation of the results of an experiment. In this view, an experiment is executed in our reality.

On the other hand, Bohr claims the necessity to specify the entire experimental setup before assigning reality to any part of it. Reality is produced by the situation that we create by the experiment. Jeeva Anandan calls this reality of quantum mechanic a **relational reality**: "Two states S_1 and S_2 are real only in relation to the interaction that they undergo with the systems they interact with." ([1]). As Anandan noticed, "relational reality" removes the antromorphic concepts from the Copenhagen interpretation.

In our view, language, logic and the reference frame of space and time result as consequences of the process to understand our measurements.

(b) Questions about future events: A clock is a subsystem of $\mathbb{W}_{opt}$ obtained by a restriction to certain sorts of measurements whose behaviour is regular with respect to time. Analogous, a translation measure is a subsystem of $\mathbb{W}_{opt}$ which is regular with respect to shifts in one space direction. With a clock and a translation measure we can therefore specify points outside the region that had been specified directly by the measurements. To forecast also the behaviour of measurements of sorts not used in these definitions, we apply the procedure of section 2. The perceptible world factually available to an observer presents in this way open horizons of space and time that broaden his actual Here and Now.

(c) The meaning of information: The simultaneous deduction of our language and the knowledge obtained in this language has important consequences. In the frame of belief theory, no conservation law holds for information ([2]). By a new message our believe in a future event may be increased or diminished and in this way our knowledge will be increased or decreased. A quantum state contains information which pertains not to the physical system but to the particular experiment we chose to perform it ([5]).

4 Explication of some quantum mechanical phenomena

4.1 The quantum Zeno effect

A description of this effect is given by Henry Stapp ([8]): If the same question is put to nature sufficiently rapidly and the initial answer is Yes, then any noise-induced diffusion, or force-induced motion, of the system away from the subensemble where the answer is "Yes" will be suppressed: the system will tend to be confined to the subensemble where the answer is "Yes".

This effect is a consequence of the way we have constructed our language. Our language is formed relative to our life world with the adjectives, which are significant for our measurements. From Principle(A), we conclude:

For being selected, an adjectives must be clearly satisfied or definitely not satisfied. A small change in the measurements will either not effect this selection or will produce a global change and effect many adjectives. Due to the minimality of the complexity of the whole description, the adjectives stabilise eachother. A changes of only a few adjectives would diminish their affiliation with the others and therefore result in an increased complexity of the whole description. The answer to a question will therefore remain stable unless a change of many adjectices occurs (that may be produced by a small change in the measurements). But such a global change affects also the formation of time.

We have selected our language in a way, that for questions formulated in our language the quantum Zeno effect holds.

4.2 The entanglement of spatially separated regions (EPR-phenomenon)

Entanglement is a consequence of the selection of our predicates and language respective to all measurements. Changes in one region may therefore affect our language and thus answers to questions put in totally different regions. This effect is well-known from daily life. If we get new experiences after the move to a new town, we will see many events from our old environment with other eyes. Our view depends now on the new and the old place and a change of this view affects our knowledge completely. It is on the other hand clear that this connection between the old and the new region can not be used for a transmission of information (No signalling property).

5 Conclusion

On the conference CASYS-2003 Gertrudis Van de Vijver summarised a fundamental experience in the sentence: "No anticipation without structure."
But which structures are combined with anticipation? To define anticipation, we have to formalise the meaning of belief. Dempster-Shafers Belief Calculus provides the simplest way to carry out this formalisation and to deduce from empirical data the structures of a world that optimally enables anticipation. In this world, that was obtained directly from measurements, we recognise the structures from quantum mechanics. Other views of the word are possible, but they depend on stronger preassumptions.

References

[1] Anandan, Jeeva (2002) Causality, Symmetries and Quantum Mechanics, Foundations of Physics Letters, Vol.15. No.5, 415-438

[2] Brukner, Caslav; Zeilinger Anton (2001) Conceptual Inadequacy of the Shannon Information in Quantum Measurements, Physical Review A, Vol.63, 022113, 1-10

[3] Cummings, Louise (2003) Formal Dialectic in Fallacy Inquiry: An Unintelligible Circumscription of Argumentative Rationality ?, Argumentation 17, 161-183

[4] Husserl, Edmund (1985) Die Phänomenologiche Methode, Reclam, Stuttgart

[5] Mehrfarin, Mohammad (2003) The Informational Nature of Quantum Mechanics: A Novel Look at the Interference Experiment, Foundations of Physics Letters, Vol.16. No.2, 127-133

[6] Sommer Hanns (2001) Voraussetzungsfreie Modellidentifikation Ed.: Haupt,P; Kersten,V.; Ulbricht,V.; Beiträge zur Modellierung und Identifikation, 1-12

[7] Sommer Hanns (1995) The Logic in Knowledge Processing Formalisms, Proceedings EUFIT95, Voll 274-278

[8] Stapp, Henry P. (2001) Quantum Theory and the Role of Mind in Nature, Foundations of Physics, Vol.31, No.10, 1465-1499

[9] Yager, Ronald R. (2001) Uninorms in Fuzzy Systems Modeling, Fuzzy Sets and Systems 122, 167-175