In: Proceedings of IEEE International Conference on System, Man and Cybernetics SMC'95.
Ref-No. 95CH3576-7, Vol. 4, pp. 3828-3832
(c) 1995 IEEE Press.

Artificial Souls: Motivated Emotional Robots

Dietrich Dörner & Katrin Hille
Universität Bamberg
D-96045 Bamberg
Germany

1. Introduction

Computer are able to play chess. They are even much better in this task than most people. Computer are also able to control a power plant. Here they are more reliable than man.

But don't ask a power plant computer to play chess. And don't ask the chess computer to control a power plant. You cannot expect them either to make suggestions for improvement. They will never look for a better way to do their jobs than the preprogrammed one. They will never change their programs automatically in case of a sudden change of the environment.

Artificial intelligence suffers from a lot of shortcomings. The systems of AI are rigid and domain specific. They are not able to adapt their behavior to the conditions of the current situation. For instance: a diagnostic system can't hurry up by looking not so detailed at its knowledge base when a quick decision is desperately needed.

An AI system does not exhibit curiosity and hence does not try to explore its environment. This is a big disadvantage. Imagine a robot putting bottles from an assembly line to a table and not realizing that there isn't a table anymore. The AI system will not notice that the environment does not present anymore the necessary conditions.

However, there is a system which doesn't share these disadvantages. This system acts autonomously and adapts its behavior to varying circumstances. It's not an AI system, it's the human being.

We are interested in action regulation of man. We tried to develop a unified theory explaining the general organization of human action regulation. — Man is a motivated emotional-cognitive system, not only a cognitive system in the sense of contemporary cognitive science. To understand the human ability to cope with complex, unknown realities it is not only necessary to explore the cognitive system of man, but also to investigate the relationships between cognition, emotion, and motivation, too.

In this article we shall demonstrate the outline of a theory unifying cognition, emotion, and motivation. We are going to describe the internal structure of the motivated emotional-cognitive system man. Additionally we shall report some preliminary tests for the soundness of this theoretical system.

2. The notion and the structure of the "Artificial Soul"

Cognitive Psychology considers psychic processes, i.e. thinking, learning, and memory-processes, as processes of information processing. There are a lot of theoretical systems which describe human thought as information processing. But how could motivation and emotion be described terms of information processing?

Motivation is based on needs. Needs indicate in our eyes set-point-deviations of certain important variables of a system, which should be kept in homeostasis, e.g. disposable energy in terms of the charge of a battery in a battery-powered robot or in terms of glucose in the blood of man. In addition to needs concerning the survival of a system (energy for instance) it seems necessary to introduce "informational" needs, e.g. needs for the exploration of the environment. So we've got curiosity. — Set-point - deviation indicators as the core of motives can easily be integrated into an information processing approach.

Emotions are different. First of all it is not very clear what the word emotions means in psychology. For some scholars emotions are something like instincts (see Plutchik, 1980). For others emotions are closely related to motivations (see Bischof, 1990). A lot of authors consider emotions to be the "ruins" of phylogenetic old, pre-intellectual forms of behavior regulation. These authors often consider emotions as unnecessary and even dangerous (see Claparède, 1928; Koestler, 1978). Other authors stress the coevolution of emotions and intellectual capacities (see e.g. Hebb, 1946), thus indicating an important role of emotions for intellectual functioning. Man is not only the most intellectual being but also, as far as we can see, the most emotional one, too.

We believe that *modulations* of cognitive processes should be considered as the core of emotions. That means emotions are not processes of their own, but the distinct forms into which cognitive processes are cast under certain conditions. To have an emotion means that the processes of behavior regulation are brought into a certain form according to the conditions of a situation.

The core of an emotion is a certain modulation of the processes preparing and controlling behavior. What does that mean? For instance: in terms of modulation of cognitive processes anger could be described as characterized by:

- High activation (which means amongst other things a high speed of information processing).
- A high "selection threshold" (i.e. a high tendency to concentrate on the actual intention and a low sensitivity for stimuli not concerning the current motivation).
- A low resolution level (i.e. a rather rough way of looking to the environment and of planning as well as decision making which does not take into account conditions for actions and side and long term effects in detail. A low level of resolution results thus in "unconditional" behavior and "overinclusive" planning.)
- A low rate of updating the image of the actual situation. (Somebody who is angry does not take into account the details of the situation and its changes.)

Angry behavior is different from "non-angry behavior" not because some independent state of "anger" is present or not, but because parameters like "selection threshold", "resolution level", "activation", and "rate of updating" have different values. The specific forms of modulation guarantee most of the time an appropriate behavior in the current situation. In a situation triggering anger it is mostly appropriate to act with verve (high activation), to persevere (high threshold of selection), to be fast

(low level of resolution), although this may mean a high degree of risk taking – and to concentrate on the source of anger instead of watching details of the environment (low rate of updating).

We understand emotion basically as a sophisticated system which modulates the cognitive processes. Emotion determines whether behavior is fast or slow, persevering or not, concentrated on the task or the environment, rough or fine. Figure 1 shows an overview on subsystems and their interactions of the system "artificial soul" (AS).

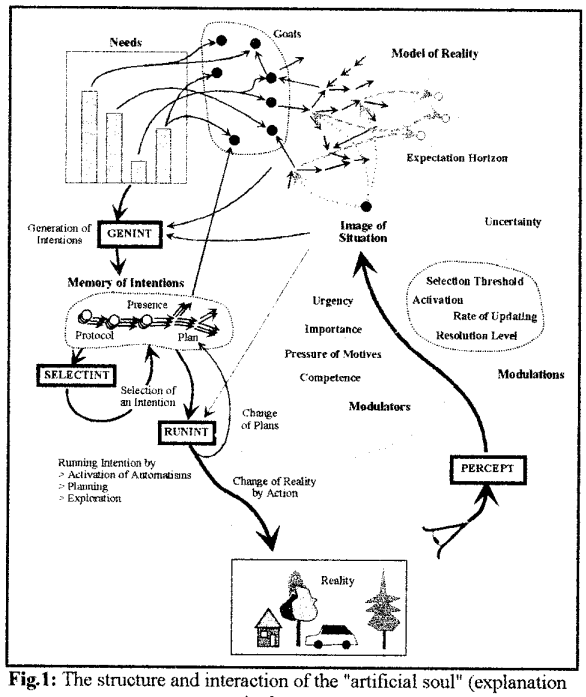

Fig.1: The structure and interaction of the "artificial soul" (explanation in the text)

On the upper left side one can see columns which represent needs of different strengths. Human beings have needs concerning survival, as for instance hunger and thirst, but also "informational" needs like affiliation and curiosity. Our ASs have different needs, too. Some of them appear periodically (e.g. the need for energy or so to speak hunger), some originate from the environmental signals (avoidance of danger), some are depending on internal conditions (e.g. curiosity which depends on events indicting the insufficiency of the knowledge-base). The needs of the AS have different strengths at different times. Needs are connected with goals. Goals are opportunities to satisfy the respective need. For instance hunger may be connected with "bread" or "apple" or "steak", thirst with "water" or "milk" or "beer".

Lets follow the line to the box GENINT. GENINT stands for *gen*eration of *int*entions. Generating an intention means choosing one goal (or several goals) which can be achieved in the given situation and an appropriate course of action. All intentions are stored in a *memory of intentions* until the underlying needs are satisfied. Most of the time this memory contains more than one intention. As it is mostly impossible to strive for several goals in the same time, it has to be decided which intention should be "run". This selection is the task of SELECTINT, which stands for *select int*ention. An intention is selected on the basis of the importance of the need, the urgency fulfilling it, and the competence of the AS for the pertinent action.

The selected intention is carried out by a system which is called RUNINT. RUNINT stands for *run int*ention. An intention can be carried out in three ways:

1. If the AS has not enough information about how to execute the intention, it *explores* different courses of action. The knowledge about the possibilities to satisfy the need will be collected by observation of the environment or by trial-and-error behavior.
2. If the AS has sufficient knowledge about the respective domain at its disposal, but doesn't know an appropriate course of action to approach the goal, the AS will engage in *planning* activities.
3. The AS will *act* if it has an appropriate plan; i.e. it will try to implement the plan.

Lets look to the right side now. The process of carrying out an intention (RUNINT) is *modulated*, i.e. the processes of exploring, planning, and implementing a plan are changed according to certain internal or external conditions. For instance: a low resolution level means rough exploration, rough (but fast) planning, and acting. A high resolution level means very careful exploration, planning, and action. The modulations themselves are triggered by "modulators". These modulators reflect on the one hand characteristics of the intentions (the urgency to carry out the actual intention, its importance, the pressure of the whole bunch of intentions in the intention memory, and the competence to run the actual intention), and on the other hand characteristics of the image of reality (i.e. the perceived uncertainty of the environment). For example: the modulation of the resolution level is influenced by

- the *importance* of the actual intention (based on the strength of the underlying need); the higher the importance of the intention, the higher the level of resolution, i.e. the more careful exploration, planning, and acting,
- the *urgency* to carry out the intention; the higher the urgency, the lower the level of resolution will be,
- the *pressure of needs* (based on the strengths of all intentions in the intention memory); the higher the activation, the lower the level of resolution.

The concept of modulators is rather close to Scherer's Stimulus Evaluation Checks (SECs) (see Scherer, 1984).

RUNINT takes into account the past of an intention. For this purpose a *protocol* of the past activities to approach the respective goal is generated. The protocol memory hinders the system for instance to repeat unsuccessful operations over and over again, as the AS will notice by looking at the protocol memory that it tried the respective action already some times.

PERCEPT stands for *percept*ion. This procedure generates an internal image of the situation by looking at the reality. PERCEPT itself is modulated by the modulators, too. For instance: according to the rate of updating, the image of the situation will be updated more or less frequently. PERCEPT yields information about events in the environment which may help or hinder the running of an intention, which may indicate opportunities to

satisfy a specific need immediately or which indicate dangers. The image of the situation is the source of the "expectation horizon". The expectation horizon is a more or less coherent picture of the future which may develop from the actual situation or may be a product of the actual behavior. An important "modulator" for an AS is the degree of branching of the expectation horizon. It means high uncertainty about the future if this degree is high. A high degree of branching will result for instance in a high rate of updating the image of the situation, as with a lot of branches in the expectation horizon an AS cannot be sure what the next moments will bring.

The motivation of an AS is represented by the needs and the goals in figure 1; the emotion by the modulator-modulation-system; cognition by the information processing in GENINT, SELECTINT, RUNINT, and PERCEPT as well as in the memory of intentions, the image of the situation, the model of reality and the expectation horizon. The thick arrows in the figure show the interactions of motivation, emotion, and cognition.

The theory exists as a computer simulation showing autonomous and adaptive behavior. The computer simulation, the AS, is able to meet the demands of a complex world:

- it is able to select one intention for execution, it decides which of the actual needs should be satisfied at the moment,
- it collects knowledge, engages in planning, and implements plans to satisfy its needs,
- it decides autonomously when to look for information, when to plan, when to use automatism,
- it adapts its behavior to conditions of its internal state and to environmental conditions.

3. The behavior of the "Artificial Soul"

We investigated the behavior of a lot of different AS in different environments. It turned out, that a "new born" AS after some "education" is able to live autonomously within its maze-like environments. Endowed with the above described system of motivation and modulation-parameters the ASs show emotional behavior. They show behavioral indicators for "fear", "hope", "anger", "resignation", "depression", and other emotions.

Imagine a dangerous situation. A signal indicates that something terrible could happen some minutes later. If one brings an AS in such a situation, one can observe distinct forms of behavior. When the AS notices the signals of danger, the modulation of the behavior will change. The activation increases: the system accelerates the speed of information processing. The selection threshold increases rapidly, too: thus the motive to avoid the dangerous situation will exclude all the other intentions from taking over the control of the behavior. — The resolution level decreases: there is no time left for *careful* exploration, planning or action. The rate of updating the image of the environment decreases: all the resources are needed to avoid the danger. Later on the rate of updating increases again: updating the image of the actual situation becomes important as approaching the dangerous situation all chances to escape the danger should be looked for.

We can observe high activation, combined with sticking not distractible to one motive. We can observe a rather rough way of planning and acting, and we can observe that the AS first stops monitoring the environment, shifting later on to eagerly looking for mouse holes. Looking at the behavior of the AS an unbiased observer will not hesitate to characterize the AS as fearful or anxious.

Figure 2 shows the protocol of the modulations in the AS in this fear and anxiety - situation. The sequence of events seems rather similar to a behavior which one would observe with human beings in a comparable situation.

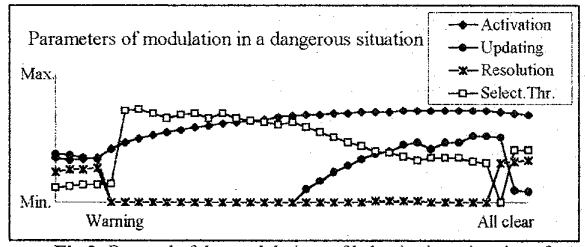

Fig.2: Protocol of the modulations of behavior in a situation of an expected, but later on not occurring dangerous event.

Not only in such a fear-and-anxiety - situation the AS behaves quite "human like". Additionally it can be shown that an AS equipped with an emotional system is "better", i.e. more able to survive in its respective environment, than a "colleague" without an emotional system.

To investigate whether this is really the case we fixed the modulations at a certain level. In this way we got "unemotional souls", the behavior of which is not any more modulated. We compared 100 of such AS without modulations with 100 emotional AS with respect to their fitness to survive. The average pressure of needs was used as a criterion for this fitness, as a low pressure of needs indicates effective coping with the challenges of the environment and a high pressure indicates that a system is not able to care for its needs.

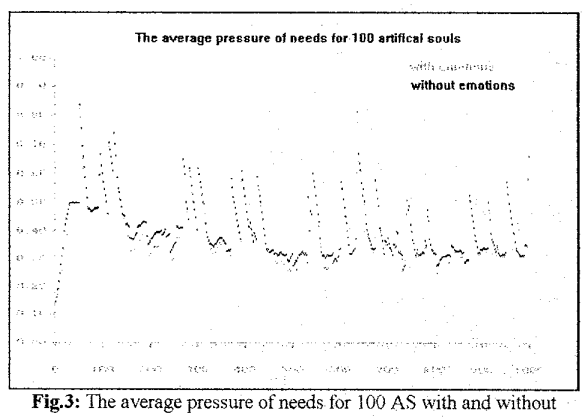

Fig.3: The average pressure of needs for 100 AS with and without emotion

Figure 3 shows the average pressure of needs for 100 ASs with and without emotion, i.e. with and without a modulation of behavior. (The peaks in the curves show the increase of need pressure in unpleasant situations, imposed by the environment.) It can be noticed that the curve of need-pressure for the AS with modulations lies below the curve of the not modulated AS. The difference of the two curves is statistically significant with a t-value of 41.7 ($p < 0.0001$). The "emotional" ASs thus show a lower pressure of needs; that means they have the more effective system of action regulation at their disposal. Emotional ASs are

fitter to survive in similar environments than unemotional ones. — This result sheds some light on the reason for the "coevolution" of cognition and emotion stated by Hebb (1946). Emotions improve the effectiveness of a cognitive system.

To find other indicators for the validity of the theoretical system behind the AS-concept we tried to create ASs with different *personalities*. Are ASs with different temperaments, for instance melancholic, phlegmatic, sanguine, and choleric ASs possible? — Eysenck characterized these well-known four temperaments of Hippocrates by values on the dimensions "unstable/stable" and "introverted/extroverted" (see Eysenck & Rachman 1965). Fig. 4 shows these two dimensions and their relation to the four temperaments of Hippocrates.

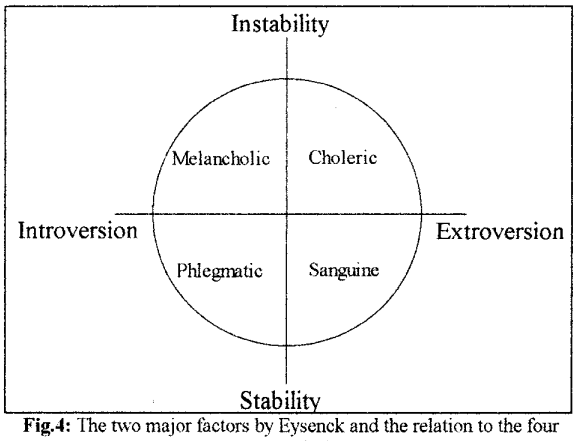

Fig.4: The two major factors by Eysenck and the relation to the four temperaments of Hippocrates

We tried to create these four temperaments: i.e. AS, which are melancholic, choleric, phlegmatic, and sanguine. For this purpose we looked for the theoretical background of the Eysenck-dimensions. Which set of parameters causes a system's stability or instability and to what constellation of parameters introversion or extroversion can be put down to?

The dimension of intro/extroversion was interpreted as dependent on a parameter "readiness to act directly to a situation". With introversion this parameter has a low value and thus introversion means a lot of thinking (in terms of our AS: exploration and planning activities) and hence less action. On the other hand, extroversion means a high readiness to react directly. The readiness to react is based on the *general competence* and the urgency of the problem. General competence is a very important parameter of the AS which is dependent on events of need-satisfaction and dissatisfaction. The satisfaction of a need adds a certain amount to general competence whereas the existence or increase of a need subtracts a certain amount from general competence. If the factor for the subtraction in the case of dissatisfaction is increased, dissatisfaction will weigh more than satisfaction. This causes an AS to "feel" incompetent and this decreases readiness to react and increases introversion. Planning plays a big role and the respective AS acts rarely and only very cautiously.

The Eysenck dimension of stability/instability was interpreted as depending on the growth rate of the needs. If under comparable conditions needs raise quicker in an AS 1 than in an AS 2, AS 1 will exhibit more behavioral variations, i.e. will show more instability. The behavior is more unstable in terms of Eysenck; the system is for instance more distractible. If the needs grow slowly, however, the behavior of the respective AS will not exhibit so much variations.

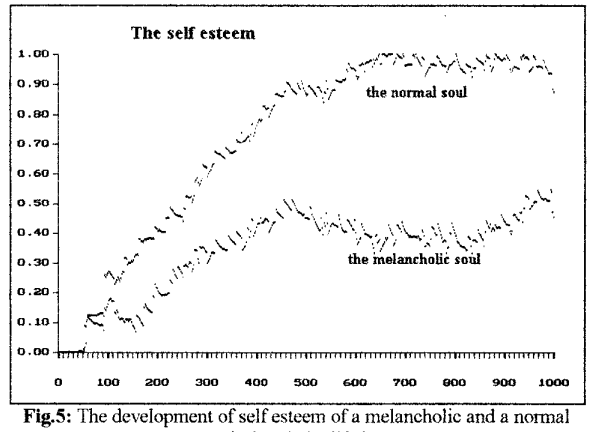

Fig.5: The development of self esteem of a melancholic and a normal AS during their "lifetime"

So we tried to create a melancholic AS by giving high values to the factor with which dissatisfaction is multiplied and low values to the growth rate of the needs. This parameter constellation resulted in a stable and introverted behavior. The general competence ("self-esteem") of a melancholic AS is much lower than that of a normal one (fig.5), although the competence based on the knowledge about the appropriate behavior ("epistemic competence") is almost the same (fig.6). Although knowing how to do it a melancholic AS does not dare to act in many cases and therefore does not develop a good "meaning" of its competence. Looking at the proportion of acting and not acting (exploring and planning), we find an AS which "thinks" a lot and does not involve in acting very much (fig.7).

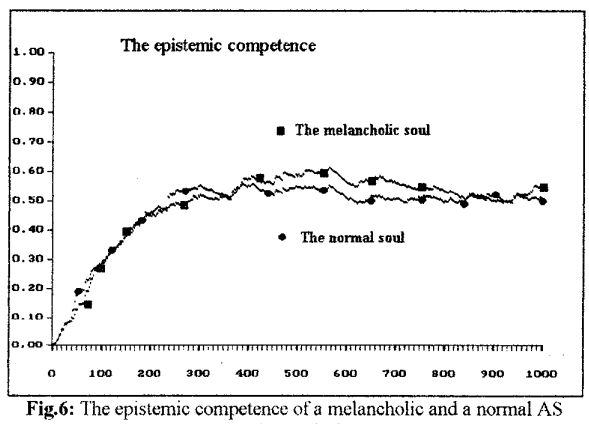

Fig.6: The epistemic competence of a melancholic and a normal AS during their life

We also created a sanguine, phlegmatic, and choleric ASs by choosing appropriate parameters for the growth rate of the needs and for the factors by which the weights of the events of satisfaction and dissatisfaction are multiplied. For sanguine ASs we increased the values of the factors for the weight of the satisfy-

ing *and* the dissatisfying events (the sanguine temperament is "thrilled to the sky and ready to die"). We created the phlegmatic and the choleric AS by varying the growth rate of needs. The impulsive choleric behavior is caused by fast increasing needs whereas the calm phlegmatic behavior can be put down to slowly growing needs. We simulated all four temperaments and tried to integrate them into Eysenck's system.

Figure 6 shows the results. The horizontal axis represents the relation of inner activity (planning) to action. The extremes of this axis represent proportions of planning to action as 1 to 3 and 1 to 4 respectively. This means that extremely introverted behavior in our case is characterized by one unit of "thinking" to three units of action whereas with extroverted behavior this proportion is 1 to 4. — The extremes of a measure of the stability/instability - scale are 0.45 and 0.35 respectively.

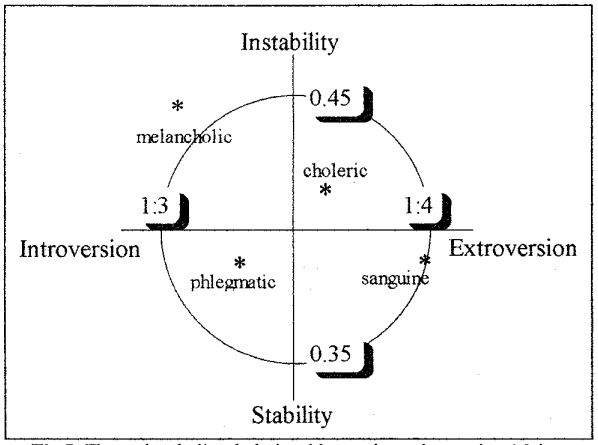

Fig.7: The melancholic, choleric, phlegmatic, and sanguine AS in Eysenck's terms of stability (pressure of needs) and extroversion (proportion planning to action)

It can be seen in fig. 7 that the melancholic AS does not do very much: for one unit of "thinking" we find approximately only three units of action. The sanguine AS acts most: for one unit of we find nearly four units of action. The highest stability is shown by the phlegmatic and sanguine AS; the lowest in the melancholic one.

4. Conclusion

The behavior of our motivated cognitive-emotional system AS seems to prove that our concept of the integration of cognitive, emotional, and motivational processes is sound and may be used to explain human behavior in complex systems and – additionally – may give some hints about what is necessary to construct autonomous robots.

We would suggest that such a robot should be supplied with

- the same or a similar organization of the processes of intention generation, intention selection, intention execution as described in fig. 1,
- the same or a similar memory organization as described in fig. 1, consisting of a model of the respective realities and a working memory, consisting of an intention memory an image of the situation, a protocol-memory and an expectation-horizon,
- a set of needs which comprise not only survival - needs like hunger (or an equivalent need for energy and matter), but informational needs, too, for instance curiosity, i.e. a need to look for new information, if the current knowledge base proves to be insufficient (i.e. incomplete or wrong),
- a modulator-modulation system (i.e. an emotional system) to adapt the form of cognitive processes to the respective internal and external conditions.

References

Bischof, N.: Emotionale Verwirrungen oder: Von den Schwierigkeiten im Umgang mit der Biologie. Wolfgang Köhler - Vorlesung auf dem 36. Kongreß der Deutschen Gesellschaft für Psychologie, S.1 - 23, 1990

Claparède, E.: Feelings and Emotions. In: M. L. Reymert (Ed.): Feelings and Emotions. The Wittenberg Symposion. Worcester / MA: Clark University Press, 1928

Dörner, D.: Wissen, Emotionen und Handlungsregulation oder die Vernunft der Gefühle. Zeitschrift für Psychologie, 2, S.167 - 202, 1993

Dörner, D.: Über die Mechanisierbarkeit der Gefühle. In: Sybille Krämer [Hrsg.]: Geist - Gehirn - künstliche Intelligenz. Berlin: de Gruyter, 1994

Eysenck, H.J. & Rachman, S. The causes and cures of neurosis. San Diego: Knapp, 1965

Hebb, D.O.: On the Nature of Fear. Psychological Review, 53, S.259 - 276, 1946

Koestler, A.: Der Mensch - Irrläufer der Evolution. Bern: Scherz, 1978

Plutchik, R.: Emotion: A Psychoevolutionary Synthesis. Harper & Row, New York, 1980

Schaub, H.: Modellierung der Handlungsorganisation. Verlag Hans Huber, Bern, 1993

Scherer, K.: On the nature and function of emotion: A component process approach. In: Scherer, K. R. & Ekman, P. (Eds.), Approaches to emotion: Hillsdale, N. Y.: Lawrence Erlbaum Association, 1984