

Modeling Emotion as an Interaction between Motivation and Modulated Cognition

Joscha Bach
joscha.bach@hu-berlin.de
*Center for Integrative Life Sciences,
Humboldt-University of Berlin, Germany*

Abstract. The cognitive architecture MicroPsi addresses the interplay between motivation and grounded cognition. Emotion is understood as a configurational state of the cognitive system, its affective dimension emerging as *a way to perform cognition* (perception, memory access, deliberation, action control etc.) rather than a module or parameter of the cognitive apparatus. We are currently adapting MicroPsi for the modeling of human behavior in problem solving test situations.

Keywords. MicroPsi, Psi theory, MicroDYN, Modulation of Cognition, Motivational System, Cognitive Architecture, Artificial General Intelligence

1. Introduction: The Cognitive Architecture MicroPsi

Within the history of Cognitive Architectures, we can discern at least the following camps:

1. *Symbolic ('classical') Models.* Newell and Simon's *Physical Symbol System Hypothesis* (1976) states that symbolic computation is both necessary and sufficient for modeling cognition, and earlier attempts at unified models of cognition took this tenet quite literally (Newell and Simon 1961; Laird, Newell, Rosenbloom 1987). While this might be theoretically true, criticism of symbolic (rule-based) approaches maintains that a purely symbolic system does not constitute a feasible *practical* approach, either because discrete symbols are technically insufficient, or because they are difficult to ground in a physical environment. This criticism gave rise to
2. *Distributed (connectionist) systems*, which focus on emergent behavior, dynamical systems and neural learning, and
3. *Embodied agents*, which focus on solving the symbol grounding problem by environmental interaction.

The two latter programs are often subsumed under the 'New AI' label, and they are vitally important: Connectionism can provide models for neural computation, for learning and perceptual processing (but will also have to explain how sub-symbolic processing gives rise to symbolic cognition, such as planning and use of natural language). Embodiment situates a system in a dynamic environment and provides content for and relevance of cognitive processes.

Unfortunately, the paradigms do not get along very well: proponents of symbolic models often ignored connectionism and symbol grounding, while connectionists frequently disregarded symbolic aspects of cognition. Most embodied modeling focuses on controlling robots instead of modeling cognition; radical proponents of embodied cognition even suggest that cognition is an *emergent* phenomenon of the interaction between an embodied nervous system and a *physical* environment (see Pfeifer and Bongard, 2006), and sometimes reject the notion of representation altogether.

The research in Cognitive Architectures has reacted to this divide by proposing neurosymbolic approaches: they combine symbolic and sub-symbolic models within a common framework (e.g., Lebière and Anderson 1993; Sun 2004), while lending themselves to situated cognition in dynamic environments.

The success of integrative, unified computational models of cognition will largely be due to the right combination of symbolic cognition (language, planning, high-level deliberation) with sub-symbolic processing (perception, analogical reasoning, neural learning and classification, memory retrieval etc.) and action regulation in a *broad architecture*.

Such a framework will have to combine *general representations* (the capability to express arbitrary relationships, up to a certain complexity) with *general learning and problem solving* (the capability to acquire and manipulate these relationships in any necessary way, up to a certain complexity), a sufficiently interesting environment to operate upon, and a *general motivational system* (which establishes the necessary goals).

MicroPsi (Bach 2003, 2009) is an attempt to add to this discussion. MicroPsi is based on the emotion and action regulation model proposed by Dietrich Dörner's Psi theory (1999, 2002), and translates it into a rigorous computational model, with grounded neurosymbolic representations. MicroPsi agents possess a motivational system based on pre-defined physiological, social and cognitive demands.

Emotions in MicroPsi are understood as

- basic affective states (moods, behavioral tendencies, simple affects), which are implemented as a modulation of cognition,
- complex emotions (specific affects directed on a motivationally relevant perceived, reflected or anticipated situation),
- communicative tools (for instance, related to the generation and interpretation of facial expressions and body language).

Currently, MicroPsi is only superficially addressing the hedonic aspects of emotion (i.e., the feelings); while for instance arousal and valence are part of the model, no self-attribution of these dimensions takes place, and the specific sensory impressions caused by the modulation of the sympathetic, parasympathetic and enteric periphery of the nervous system are ignored altogether. Also, while the principle of emergence of emotional states is relatively well-covered in the MicroPsi context, we have not yet attempted to deliver a taxonomy of human emotions (with their respective implementation), nor a grounding in other areas of psychology, for instance in the "big five" of personality structure (see, for instance, Digman 1990).

2. Motivation in MicroPsi: Generating Relevance

In my view, emotion cannot be modeled as an isolated component—it is always part of a larger cognitive architecture, including a motivational system that may attach relevance to cognitive content. Desires and fears, affective reflexes and mood changes correspond to *needs*, such as environmental exploration, identification and avoidance of danger, and the attainment of food, shelter, cooperation, procreation, and intellectual growth. Since the best way to satisfy the individual needs varies with the environment, the motivational system is not aligned with particular *goal situations*, but with the needs themselves, through a set of *drives*.

Let us call events that satisfy a need of the system a *goal*, or an *appetitive event*, and one that frustrates a need an *aversive event* (for instance, a failure or an accident). Since goals and aversive events are given by an open environment, they can not be part of the definition of the architecture, and the architecture must specify a set of drives according to the needs of the system. Drives are indicated to the system as *urges*, as signals that make a need apparent. An example of a need would be nutrition, which relates to a drive for seeking out food. On the cognitive level of the system, the activity of the drive is indicated as *hunger*.

The connection between urges and events is established by *reinforcement learning*. In our example, that connection will have to establish a representational link between the indicator for food and a *consumptive action* (i.e., the act of ingesting food), which in turn must refer to an environmental situation that made the food available. Whenever the urge for food becomes active in the future, the system may use the link to retrieve the environmental situation from memory and establish it as a goal.

Dörner's Psi theory follows this paradigm. Unlike high-level descriptions of motivation as they are more common in psychology, such as the one by Maslov (1987) or Kuhl (2001), the theory is rigorous enough to be implemented as a computational model, and unlike narrow, physiological models (such as Tyrell, 1993), it also addresses cognitive and social behavior. I have adapted the Psi theory for implementing computer simulations of virtual agents (Bach 2007), and I will identify the core components of the motivational system.

2.1. Needs

All urges of the agent stem from a fixed and finite number of 'hard-wired' needs, implemented as parameters that tend to deviate from a target value. Because the agent strives to maintain the target value by pursuing suitable behaviors, its activity can be described as an attempt to maintain a *dynamic homeostasis*.

Currently, the agent model of the Psi theory suggests several "physiological" needs (fuel, water, intactness), two "cognitive" needs (certainty, competence) and a social need (affiliation).

All behavior of Psi agents is directed towards a goal situation that is characterized by a *consumptive action* satisfying one of the needs. In addition to what the physical (or virtual) embodiment of the agent dictates, there are cognitive needs that direct the agents towards exploration and the avoidance of needless repetition.

The needs of the agent are weighted against each other, so differences in importance can be represented.

2.1.1. *Physiological needs*

Fuel and water: In our simulations, water and fuel are used whenever an agent executed an action, especially locomotion. Certain areas of the environment caused the agent to loose water quickly, which associated them with additional negative reinforcement signals.

Intactness: Environmental hazards may damage the body of the agent, creating an increased intactness need and thus leading to negative reinforcement signals (akin to *pain*). If damaged, the agent may look for opportunities for repair, which in turn increase intactness.

These simple needs can be extended at will, for instance by needs for shelter, for rest, for exercise, for certain types of nutrients etc.

2.1.2. *Cognitive needs*

Certainty: To direct agents towards the exploration of unknown objects and affairs, they possess an urge specifically for the reduction of uncertainty in their assessment of situations, knowledge about objects and processes and in their expectations. Because the need for certainty is implemented similar to the physiological urges, the agent reacts to uncertainty just as it would to pain signals and will display a tendency to remove this condition. This is done by triggering explorative behavior.

Events leading to an urge for uncertainty reduction include:

- the agent meets unknown objects or events,
- for the recognized elements, there is no known connection to behavior—the agent has no knowledge what to do with them,
- there are problems to perceive the current situation at all,
- there has been a breach of expectations; some event has turned out differently as anticipated,
- over-complexity: the situation changes faster than the perceptual process can handle,
- the anticipated chain of events is either too short or branches too much. Both conditions make predictions difficult.

In each case, the uncertainty signal is weighted according to the relation to the appetitive or aversive relevance of the object of uncertainty.

The urge for certainty may be satisfied by “certainty events”—the opposite of uncertainty events:

- the complete identification of objects and scenes,
- complete embedding of recognized elements into agent behaviors,
- fulfilled expectations (even negative ones)
- a long and non-branching chain of expected events

Like all urge-satisfying events, certainty events create a positive reinforcement signal and reduce the respective need. Because the agent may anticipate the reward signals from successful uncertainty reduction, it can actively look for new uncertainties to explore (“diversive exploration”).

Competence: When choosing an action, Psi agents weight the strength of the corresponding urge against the chance of success. The measure for the chance of

success to satisfy a given urge using a known behavior program is called “specific competence”. If the agent has no knowledge on how to satisfy an urge, it has to resort to “general competence” as an estimate. Thus, general competence amounts to something like self-confidence of the agent, and it is an urge on its own. (Specific competencies are not urges.)

The general competence of the agent reflects its ability to overcome obstacles, which can be recognized as being sources of negative reinforcement signals, and to do that efficiently, which is represented by positive reinforcement signals. Thus, the general competence of an agent is estimated as a floating average over the reinforcement signals and the inverted displeasure signals. The general competence is a heuristic on how well the agent expects to perform in unknown situations.

As in the case of uncertainty, the agent learns to anticipate the positive reinforcement signals resulting from satisfying the competence urge. A main source of competence is the reduction of uncertainty. As a result, the agent actively aims for problems that allow to gain competence, but avoids overly demanding situations to escape the frustration of its competence urge. Ideally, this leads the agent into an environment of medium difficulty (measured by its current abilities to overcome obstacles).

Aesthetics: Environmental situations and relationships can be represented in infinitely many ways. Here ‘aesthetics’ corresponds to a need for improving representations, mainly by increasing their sparseness, while maintaining or increasing their descriptive qualities.

2.1.3. Social needs

Affiliation: Because the explorative and physiological desires of our agents are not sufficient to make them interested in each other, they have a need for positive social signals, so-called ‘*legitimacy signals*’. With a legitimacy signal (or *l-signal* for short), agents may signal each other “okayness” with regard to the social group. Legitimacy signals are an expression of the sender’s belief in the social acceptability of the receiver. The need for l-signals needs frequent replenishment and thus amounts to an urge to affiliate with other agents. Agents can send l-signals to reward each other for successful cooperation.

Anti-l-signals are the counterpart of l-signals. An anti-l-signal (which basically amounts to a frown) ‘punishes’ an agent by depleting its legitimacy reservoir.

Agents may also be extended by *internal l-signals*, which measure the conformance to internalized social norms.

Supplicative signals are ‘pleas for help’, i.e. promises to reward a cooperative action with l-signals or likewise cooperation in the future. Supplicative signals work like a specific kind of anti-l-signals, because they increase the legitimacy urge of the addressee when not answered. At the same time, they lead to (external and internal) l-signals when help is given. They can thus be used to trigger *altruistic behavior*.

The need for l-signals should adapt to the particular environment of the agent, and may also vary strongly between agents, thus creating a wide range of types of social behavior. By making the receivable amount of l-signals dependent of the priming towards particular other agents, Psi agents might be induced to display ‘*jealous behavior*’.

Social needs can be extended by romantic and sexual needs. However, there is no explicit need for social power, because the model already captures social power as a specific need for competence—the competence to satisfy social needs.

Even though the affiliation model is still fragmentary, we found that it provides a good handle on the agents during experiments. The experimenter can attempt to induce the agents to actions simply by the prospect of a smile or frown, which is sometimes a good alternative to a more solid reward or punishment.

2.2. *Appetence and aversion*

In order for an urge to have an effect on the behavior on the agent, it does not matter whether it *really* has an effect on its (physical or simulated) body, but that it is represented in the proper way within the cognitive system. Whenever the agent performs an action or is subjected to an event that reduces one of its urges, a reinforcement signal with a strength that is proportional to this reduction is created by the agent's "pleasure center". The naming of the "pleasure" and "displeasure centers" does not necessarily imply that the agent experiences something like pleasure or displeasure. The name refers to the fact that—like in humans—their purpose lies in signaling the reflexive evaluation of positive or harmful effects according to physiological, cognitive or social needs. (*Experiencing* these signals would require an observation of these signals at certain levels of the perceptual system of the agent.)

2.3. *Motives*

A motive consists of an urge (that is, the value of an indicator for a need) and a goal that has been associated to this indicator. The goal is a situation schema characterized by an action or event that has successfully reduced the urge in the past, and the goal situation tends to be the end element of a behavior program. The situations leading to the goal situation—that is, earlier stages in the connected occurrence schema or behavior program—might become intermediate goals. To turn this sequence into an instance that may initiate a behavior, orient it towards a goal and keep it active, we need to add a connection to the pleasure/displeasure system. The result is a *motivator* and consists of:

- a need sensor, connected to the pleasure/displeasure system in such a way, that an increase in the deviation of the need from the target value creates a displeasure signal, and a decrease results in a pleasure signal. This reinforcement signal should be proportional to the strength of the increment or decrement.
- optionally, a feedback loop that attempts to normalize the need automatically
- an urge indicator that becomes active if there is no way of automatically adjusting the need to its target value. The urge should be proportional to the need.
- an associator (part of the pleasure/displeasure system) that creates a connection between the urge indicator and an episodic schema/behavior program, specifically to the aversive or appetitive goal situation. The strength of the connection should be proportional to the pleasure/displeasure signal. Note that usually, an urge gets connected with more than one goal over time, since there are often many ways to satisfy or increase a particular urge.

3. The Dynamics of Modulation: the Basis of Affect

In the course of the action selection and execution, MicroPsi agents are modulated by several parameters: The agent's *activation* or *arousal* (which resembles the *ascending reticular activation system* in humans) determines the action-readiness of an agent. It is proportional to the current strength of the urge signals. The perceptual and memory processes are influenced by the agent's *resolution level*, which is inversely related to the activation. A high resolution level increases the number of features examined during perception and memory retrieval, at the cost of processing speed and resulting ambiguity. The *selection threshold* determines how easily the agent switches between conflicting intentions and thus avoids goal oscillation, and the *sampling rate* or *securing threshold* controls the frequency of reflective and orientation behaviors. The values of the modulators of an agent at a given time, together with the status of the urges, define a cognitive configuration, a setup that may be interpreted as an *emergent emotional state*, located in a multi-dimensional space.

One of the first attempts to treat emotion as a continuous space was made by Wilhelm Wundt (1910). According to Wundt, every emotional state is characterized by three components that can be organized into orthogonal dimensions. The first dimension ranges from pleasure to displeasure, the second from arousal to calmness, and the last one from tension to relaxation (figure 1), that is, every emotional state can be evaluated with respect to its positive or negative content, its stressfulness, and the strength it exhibits. Thus, an emotion may be pleasurable, intense and calm at the same time, but not pleasurable and displeasurable at once. Wundt's model has been re-invented by Charles Osgood in 1957, with an *evaluation* dimension (for pleasure/displeasure), *arousal*, and *potency* (for the strength of the emotion) (Osgood et al. 1957), and re-discovered by Ertel (1965) as *valence*, *arousal*, and *potency*.

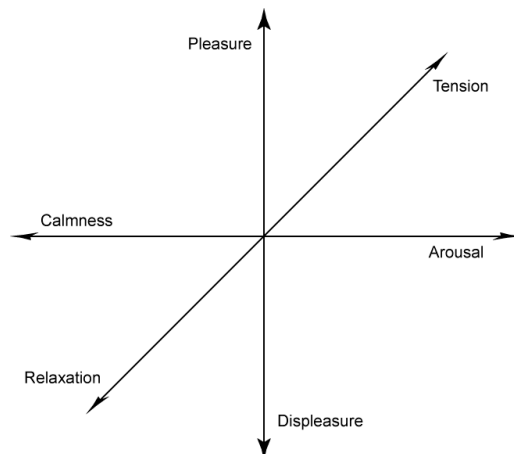


Figure 3: Dimensions of Wundt's emotional space (see Wundt 1910)

Wundt's model does not capture the social aspects of emotion, so it has been sometimes amended to include extraversion/introversion, apprehension/disgust and so

on, for instance by Traxel and Heide, who added *submission/dominance* as the third dimension to a *valence/arousal* model (Traxel and Heide 1961).

Note that *arousal*, *valence* and *introversion* are themselves not emotions, but mental configuration parameters that are much closer to the physiological level than actual emotions – we could call them *proto-emotions*. Emotions are areas within the space spanned by the proto-emotional dimensions.

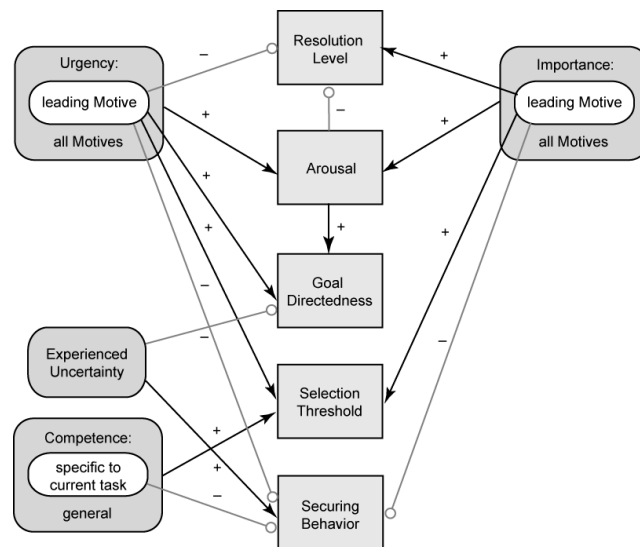


Figure 2: Dimensions of emotional modulation in MicroPsi

The modulators currently defined in MicroPsi are adapted from Katrin Hille's doctoral thesis on emotional dynamics (1997). They span a continuous six-dimensional emotional space (figure 2), described with the aforementioned proto-emotional dimensions: *arousal* (which is equivalent to Wundt's *arousal* dimension), *resolution level*, *dominance* of the leading motive (usually called *selection threshold*), the level of *background checks* (the rate of the securing behavior), and the level of *goal-directed behavior*. (The latter two may loosely correspond to Wundt's *tension* dimension). The sixth dimension is the *valence*, i.e. the signals supplied by the pleasure/displeasure system. The dimensions are not completely orthogonal to each other (resolution is mainly inversely related to arousal, and goal orientedness is partially dependent on arousal as well).

The six-dimensional model is not exhaustive; especially when looking at social emotions, at least the demands for *affiliation* (external legitimacy signals) and 'honor' (internal legitimacy, ethical conformance), which are *motivational dimensions* like *competence* and *uncertainty reduction*, would need to be added.

It would be worthwhile to discuss the implementation of complex emotions (i.e., directed affects), to address the expression of emotion, or to properly contrast the modulation space model of emotion used in MicroPsi with other approaches. MicroPsi is a functionalist, system-level model, as opposed to agent-level models (as exemplified in the classic OCC model, 1988), or non-functionalist descriptive approaches (for

instance, appraisal models, e.g. Scherer 1993). In the available space, only a short introduction to MicroPsi can be given, and I would like to refer the reader to other publications on these topics (Bach 2006, 2007).

4. Outlook

MicroPsi is both a theoretical framework for a cognitive architecture, and a software development effort to implement computational models of behavior. MicroPsi is available as a downloadable multi-platform plugin for the Eclipse IDE, and has been applied to Artificial Life simulations, category learning, and as a robot control architecture.

After a period of AI research for industry application, during which the development of MicroPsi was dormant, we have found a new home for our research at the interdisciplinary Center for Integrative Lifesciences at the Humboldt University of Berlin. We are currently working to extend MicroPsi on two fronts, both as an AI architecture, and as a psychological model of human behavior.

Within the OpenPsi project, together with Ben Goertzel's Artificial Intelligence group at the Polytechnical University of Hong Kong, we export MicroPsi's model of emotion and motivation into the AI architecture OpenCog (Hart and Goertzel 2008, Goertzel 2009), with the goal of making it available for creating believable agents, especially in computer games.

Together with Joachim Funke's department for psychology at the University of Heidelberg, we aim to adapt MicroPsi to model the decision making strategies of specific groups of psychiatric patients in problem solving games, with the goal of obtaining diagnostically applicable metrics (MicroDYN: Greiff and Funke 2009).

Acknowledgments

This work is being supported by a fellowship of the Center for Integrated Life Sciences (CILS) at the Humboldt University of Berlin.

References

- Bach, J. (2003). The MicroPsi Agent Architecture Proceedings of ICCM-5, International Conference on Cognitive Modeling, Bamberg, Germany, 15-20
- Bach, J. (2006): MicroPsi: A cognitive modeling toolkit coming of age. In Proceedings of 7th International Conference on Cognitive Modeling (ICCM06): 20-25
- Bach, J. (2007): Motivated, Emotional Agents in the MicroPsi Framework, in Proceedings of 8th European Conference on Cognitive Science, Delphi, Greece
- Bach, J. (2009). Principles of Synthetic Intelligence. Psi, an architecture of motivated cognition. Oxford University Press.
- Digman, J.M. (1990). Personality structure: Emergence of the five-factor model. Annual Review of Psychology 41: 417-440
- Dörner, D. (1999). *Bauplan für eine Seele*. Reinbeck
- Dörner, D., Bartl, C., Detje, F., Gerdes, J., Halcour, (2002). *Die Mechanik des Seelenwagens. Handlungsregulation*. Verlag Hans Huber, Bern

- Ertel, S. (1965). Standardisierung eines Eindrucksdifferentials. *Zeitschrift für experimentelle und angewandte Psychologie*, 12, pp 22-58
- Goertzel, B. (2009). Opencog prime: A cognitive synergy based architecture for embodied artificial general intelligence. In *Proceedings of ICCI-09, Hong Kong, 2009*
- Greiff, S., Funke, J. (2009). Measuring complex problem solving - The MicroDYN approach. In F. Scheuermann (Ed.), *The transition to computer-based assessment - Lessons learned from large-scale surveys and implications for testing*. Luxembourg: Office for Official Publications of the European Communities.
- Hart D., Goertzel, B. (2008). OpenCog: A Software Framework for Integrative Artificial General Intelligence. In *Proceedings of AGI2008*. pp.468~472
- Hille, K. (1997): Die „künstliche Seele“. Analyse einer Theorie. [The artificial soul. Analysis of a theory.] Wiesbaden: Deutscher Universitäts-Verlag
- Kuhl, J. (2001): *Motivation und Persönlichkeit: Interaktionen psychischer Systeme*. Göttingen: Hogrefe
- Maslow, A., Frager, R., Fadiman, J. (1987): *Motivation and Personality*. (3rd edition.) Boston: Addison-Wesley
- Laird, J. E., Newell, A., Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, 33(1), 1-64
- Lebiere, C., Anderson, J. R. (1993): A Connectionist Implementation of the ACT-R Production System. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, pp. 635-640
- Newell, A., Simon, H. A. (1961): GPS, a program that simulates human thought, in: E. Feigenbaum and J. Feldmann (eds.) (1995): *Computers and Thought*
- Newell, A., Simon, H. A. (1976): Computer Science as Empirical Inquiry: Symbols and Search, *Communications of the ACM*, vol. 19 (March 1976): 113—126
- Ortony, A., Clore, G. L., Collins, A. (1988): *The Cognitive Structure of Emotions*. Cambridge, England: Cambridge University Press
- Osgood, C. E., Suci, G. J. Tannenbaum, P. H. (1957): *The measurement of meaning*. Urbana: University of Illinois Press
- Pfeifer, R., Bongard, J. (2006): *How the body shapes the way we think*. MIT Press
- Scherer, K. (1993): Studying the Emotion-Antecedent Appraisal Process: An Expert System Approach. In: *Cognition and Emotion*, 7 (3/4), p. 325-355
- Sun, R. (2004). *The Clarion Cognitive Architecture: Extending Cognitive Modeling to Social Simulation*
- Traxel, W. (1960): Die Möglichkeit einer objektiven Messung der Stärke von Gefühlen. *Psychol. Forschung*, 75-90
- Traxel, W., Heide, H. J. (1961): Dimensionen der Gefühle. *Psychol. Forschung*, 179-204
- Tyrell, T. (1993): *Computational Mechanism for Action Selection*, PhD Thesis, University of Edinburgh
- Wundt, W. (1910): Gefühlselemente des Seelenlebens. In: *Grundzüge der physiologischen Psychologie II*. Leipzig: Engelmann