A Framework for Emergent Emotions, Based on Motivation and Cognitive Modulators

Joscha Bach

Berlin School of Mind and Brain, Humboldt University of Berlin
Unter den Linden 6, 10199 Berlin, Germany
E-mail: joscha.bach@hu-berlin.de

Abstract: While traditional appraisal models have been successful tools for describing and formalizing the behavior of emotional agents, they have little to say about the functional realization of affect and emotion within the cognitive processing of these agents. The cognitive architecture MicroPsi addresses emotion and motivation by defining pre-requisites over which affective dynamics and goal-seeking emerge. Here, these pre-requisites are explained in detail, along with a possible approach of using them to model personality traits.

Keywords: emergent emotions, directed affects, autonomy, motivation, MicroPsi, cognitive modulation

Introduction

Emotion and affect are intrinsic to our cognition, and any attempt at a detailed understanding of the human mind will require attention to this domain (Sloman, 1981; Lisetti & Gmytrasiewicz, 2002). While models of emotion have immediate applications, for instance in human computer interaction and user modeling, their main significance might lie deeper: the question of how it is possible that a mind, a biologically implemented information processing machine, is able to feel, to undergo emotional episodes, to turn into a self-reflecting and social agent has remained a dazzling issue to the philosophy of mind and cognitive science in general, and an adequate functional model of emotions will be an important part of the answer.

Computational modeling of emotion and affect has seen a wide variety of different approaches, which I will not review here. (Authoritative summaries on the state of the art may be found elsewhere, for instance in Gratch, Marsella, & Petta, 2011; for a look at its history consult Hudlicka & Fellous, 1996; Gratch & Marsella, 2005.) The nature of these approaches has been largely determined by applications, for instance for behavior modeling, for supporting communication with artificial systems, and for social simulations. Such applications favor externalist, descriptive models of emotional agents, for instance in *belief/desire/intention* frameworks (*BDI:* Bratman, 1987). Conversely, if the goal is an understanding of cognitive behaviors, self-assessment of agents, the mechanisms of filtering and biasing in memory access, perception and action control, and the relationship between emotion and motivation, we require internalist, functional models.

Externalist models arguably dominate today's research in synthetic emotions, with a focus on the very successful family of *appraisal theories of emotion* (see Roseman, 1991; Lazarus, 1991; Ellsworth & Scherer, 2003). Appraisals reflect assessments of external and internal stimuli of an agent, and they give rise behavioral and dispositional consequences. The intensity and range of affects and emotions is subject to individual variance (Russel, 1995), and their directedness is the result of adaptive learning, but the dimensionality, general expression and cognitive structure of emotions is largely invariant (Ekman & Friesen, 1971; Izard, 1994). For example, while a person might learn in what situations fear is appropriate or inappropriate, the ability to perceive fear itself is not acquired, rather, it stems from the way its organism is equipped to react to certain external or internal stimuli. Thus, it makes sense to develop general taxonomies of emotional states. The well-known *Ortony-Clore-Collins model* (*OCC:* Ortony, Clore & Collins, 1988) represents a high-level classification of these assessments: It treats emotions as *valenced reactions* to the consequences of events, to the actions of agents, or to aspects of objects, by distinguishing whether those situations and actions are desirable or undesirable, happen to oneself or another agent, are manifest or projected and so on (figure 1). The OCC model elegantly captures the difference between social emotions (the appraisal of actions for oneself and others) and event-based emotions like hope or relief. Even though it is not exhaustive (in its original form, it does not account for all high-level emotions like jealousy or envy), it scales easily by adding additional appraisal conditions.

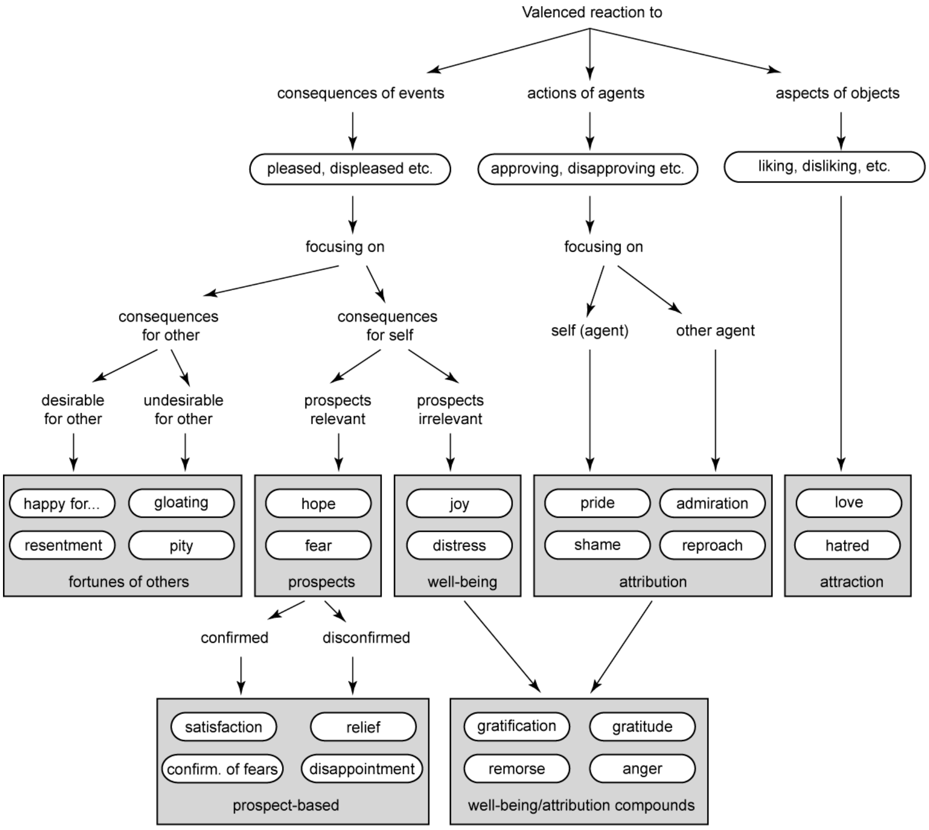

Figure 1: Taxonomy of higher-level emotions (after Ortony, Clore & Collins, 1988, p. 19)

Based on the OCC model, every emotion can be specified in a formal language, with real-valued threshold parameters to specify intervals in a weight-matrix to describe

- For *events*: their desirability for the agent itself, their desirability for others, their deservingness, their liking, the likelihood of their occurrence, the related effort, and whether they are realized,
- For *agents*: their praiseworthiness, their cognitive relevance, the deviation of expectations,
- For *objects*: their appeal and their familiarity.

Because these aspects form a clear ontology, their adoption for computational models is straightforward, and has influenced a wide variety of applications, from the creation of believable agents for computer games to economic simulation.

On the other hand, externally observable behavioral consequences of emotion cover only a small area of their function. For instance, emotions and affective states also give internal feedback on an agent's performance and aid in reflection, structure social interaction, enhance communication. Beyond a valenced feedback for learning and interaction with the environment, they also prime memory retrieval by providing associative cues, bias perception and filter the access to mental content according to situational dynamics, for instance by triggering faster or more sophisticated processing, deeper or wider associativity, stronger or weaker goal adherence etc., in short: they *modulate* the cognitive processing itself. Capturing these aspects of emotion requires addressing its functional realization within a cognitive architecture. Beyond valenced reactions and their classification, such a model must describe the emergence of emotion and affect in the first place, by realizing them functionally.

Here, I will detail a possible approach for such a functional realization, using the cognitive architecture *MicroPsi* (Bach, 2003, 2007, 2009, 2011). MicroPsi focuses on capturing autonomous behavior and the grounding of neuro-symbolic representations, but also features motivation and emotion as integral parts. Originally based on a model from theoretical psychology, the *Psi theory* by Dietrich Dörner (1999, 2002), it treats emotion and motivation as distinct, but related facets of cognition. This contribution will describe the underlying model (which has been partially done elsewhere: see Bach, 2009), and explain its application for the modeling of personality traits.

Setting up the conceptual frame

Defining emotions in their broad sense is notoriously difficult and prone to misunderstandings; so instead of a general definition, I would like to use the following, more narrow terminology.

The *motivational system* of an agent is based on a set of systemic needs, or *demands*, which are represented as *drives*. A drive manifests as an *urge signal* that influences behavior and learning of the agent according to its needs. On the lowest level, these influences are *modulators* of cognition, such as mechanisms for realizing continuous changes in *arousal*, and continuous evaluations of *valence*. Here, the arousal is a cognitive parameter that controls the general activation and action readiness of an individual, and depends on the urgency of its perceived needs of that agent. The valence is a reinforcement signal reflecting changes in those needs: a rapid decrease of a demand amounts to a positive valence (‘pleasure’), and a sudden increase in a demand results in a negative valence (‘distress’).

This lowest level already captures some affective phenomena, like *affective reflexes* (especially *startling*, an alarm reaction characterized by a sudden increase of arousal due to a severe unexpected perceptual mismatch), and the valenced states caused by the pleasure signals (*joy*, *distress*, *anxiety*).

The cognitive modulators define several axes of a configuration space of cognition. Undirected *moods* (euphoria, depression etc.) correspond to regions of that configuration space. More generally, *cognitive configurations*, or modes of cognition, form the secondary level of the description: the *affective state*.

Affects are often directed upon an object, which is determined by its motivational relevance. This gives rise to what I would like to term a *higher-level emotion*. The object of an affective configuration may either be a *state* (as in jealousy, pity or pride), or it could be a motivationally relevant *process* (i.e. a belief change, as in relief or disappointment).

Thus, we will distinguish between *demands*, *urges*, *modulators*, *affective states* and *directed (higher-level) emotions*. Demands and urges are part of the motivational system in our terminology, and while they may give rise to emotion, they are conceptually distinct. Simply put: Motivation determines what is to be done, emotion shapes how it is being done.

I suggest that for capturing the emergence of emotions, the motivational and modulatory processes are necessary and sufficient: the emotions themselves are not causal structures (parameters or modules) that are to be represented at the architectural level. Instead, emotions are best understood as perceptual gestalts (Castelfranchi & Miceli, 2009). We arrive at emotions by perceiving aspects of motivation and modulation internally, or behavioral consequences externally, and categorize these cognitive configurations into precisely those categories described by the externalist models, such as the OCC model.

To account for the necessary aspects of motivation and modulation, a cognitive architecture must capture the following components (see e.g. Diener, 1999):

- The subjective experience of emotions (how it feels to be in an emotional state). This involves valence, proprioception and the reflection of qualities of cognitive behaviors, such as rumination, goal-directedness, action-readiness and so on.
- The cognitive correlates of the physiological mechanisms (neural, neurochemical, feedback from muscular activation etc.) that facilitate emotion, such as behavior regulation, attentional focusing, task switching etc.
- Emotional expression (facial expression, body posture, movement patterns, modulation of voice and breathing etc.).
- The cognitive evaluation of stimuli and the agent's own behavior.
- Changes in behavioral and perceptual dispositions.

These components describe a feature space over which culturally defined emotion categories may be attributed.

This approach to modeling emotion as emergent perceptual gestalts is not without alternatives. Other treatments include:

- Emotions may be seen as explicit states. The emotional agent has a number of states it can adopt, possibly with varying intensity, and a set of state transition functions. These states may be used to parameterize the modules of behavior, perception, deliberation and so on.
- Emotions can be defined as direct functions of beliefs and desires of an agent (Reisenzein, 2001).
- Modeling emotions by connecting them directly to stimuli, assessments or urges (like hunger or social needs) of the agent. (A similar approach has been suggested by Frijda, 1986.)

- Disassembling emotions into compounds (sub-emotions, basic emotions), and modeling their co-occurrence. Some suggestions for suitable sets of primary emotions and/or emotion determinants have been made by some emotion psychologists (for instance Plutchik, 1994).

Conversely, our approach is going to capture emotions implicitly, because we do not see them as natural kinds. Instead, we identify the parameters that modify the agent's cognitive behavior and are thus the correlates of the emotions. The manipulation of these parameters leads to the emergence of affective states, and the combination of affective states with motivationally relevant mental content amounts to directed emotions.

Architectural requirements

If we describe emotions as aspects of more basic cognitive processes, we will need to make these processes explicit by specifying requirements to a more general architecture of cognition. Such an architecture will need operations that can be objects of motives, representations that can capture them, and processes that can be modulated (figure 2). Consequently, we require:

- A set of *urges* that signal demands of the agent.
- A *selection mechanism* that promotes the satisfaction of one of the urges to an *intention* (an active motive).
- An *action selection/planning mechanism* that chooses actions to reach the goal associated with satisfying the urge.
- An *associative memory*, which can be *primed* (pre-activated or biased for) by active motives.
- *Action execution* mechanisms that actually perform the chosen actions.
- A set of *modulators* that modify the access to memory content, and the way perception, action selection and action execution work.
- A *reinforcement learning* mechanism that creates associations between urges, goal situations and aversive events, based on the effect that encountered situations have on the demands.

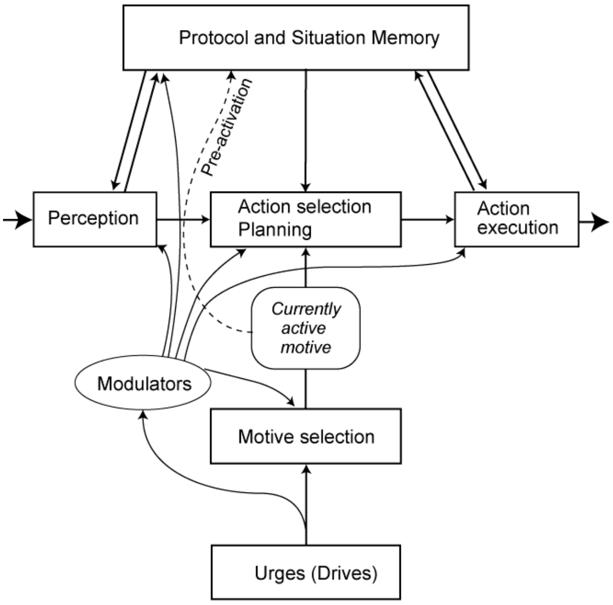

Figure 2: Minimal architectural requirements for emergent emotions (see Bach, 2009, p. 68)

Depending on the set of modulators, we may need additional requirements with respect to the memory, beyond motivational priming. For instance, if we equip the memory with a spreading activation paradigm, we can modulate the width and depth of activation spreading, thereby accounting for a variable level or resolution for mental representations. Many affective states are characterized by variances in the resolution of mental content; especially: a high arousal corresponds to a lower level of detail (Cole, Michel & O'Donnell Teti, 1994).

More generally, the memory needs to provide the following structures (figure 3):

- A *situation image* (which holds a model of the current state of the world and the agent itself).
- A *long-term memory*, derived from a protocol of situation images. Here, the agent maintains a long-term *self-model*, a *declarative memory* (stable object and category abstractions), and an *episodic memory* (including procedural/skill knowledge).
- An *inner stage*, to maintain hypothetical, anticipated and counterfactual representations, especially expectations, intended situations (goals) and plans.

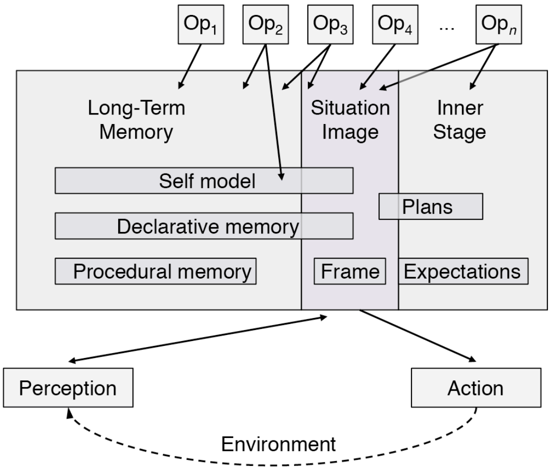

Figure 3: Basic memory structures for an emotional agent

These components are required to capture the range of object-related emotions involving situation assessment, reflection of past events (such as regret) and the anticipation of future events (dread, hope etc.). The operations that manage and maintain the agent's memory content (such as abstraction, categorization, retrieval, matching, deliberation, planning, analogical reasoning, construction, extrapolation, interpolation etc.) can be subjected to modulation and thereby provide the configurations that correspond to affective states.

These basic requirements sketch the frame for a large variety of possible cognitive architectures that can be used to model emergent emotions; for instance, it is possible to identify them to a large degree in Aaron Sloman's *CogAff* framework (Sloman, Chrisley & Scheutz, 2005), or Ron Sun's *CLARION* (Sun, 2004, 2005). In the following, let us look at the realization of modulation, motivation and higher-level emotion as they are realized in MicroPsi.

Cognitive modulators, and their interpretation in MicroPsi

In section 2, we suggested two basic parameters that might be used to determine the space of possible affects: valence (pleasure/displeasure) and arousal. If we add a third dimension, *tension/relaxation*, we arrive at Wilhelm Wundt's historical emotion space model (figure 4). According to Wundt (and later on Woodworth, 1938; Osgood, 1957; Ertel, 1965) each emotion can be analyzed according to these three orthogonal aspects, i.e. it is characterized by its pleasurableness, its stressfulness, and its intensity. Thus, an emotion may be pleasurable, intense and calm at the same time, but not pleasurable and displeasurable at once. Since Wundt's model does not capture the social aspects of emotion, it has been sometimes amended to include extraversion/introversion, apprehension/disgust and so on, for instance by Traxel and Heide (1961) and Mehrabian and Russell (1980), who added *submission/dominance* as the third dimension to a *valence/arousal* model, and Schlosberg (1954), who called the third dimension *acceptance/rejection*.

Pleasure/arousal/dominance models (PAD) have been adopted for numerous agent implementations, often in combination with the OCC model (e.g., WASABI: Becker-Asano, 2008; FAtiMA: Dias & Paiva, 2005; Peña, Peña & Ossowski, 2012).

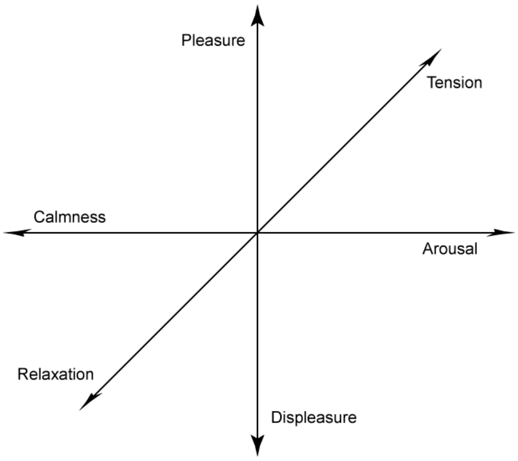

Figure 4: Dimensions of Wundt's emotional space (after Wundt 1910)

Obviously, *arousal*, *valence*, *stress,* and *dominance* are not emotions, affects or moods by themselves: they are cognitive modulators, closely aligned with physiological parameters. However, characteristic co-occurrences of modulator settings configure the cognition of an individual in specific ways, which we can classify as affects. Affective states are regions in the space spanned by the cognitive modulators.

MicroPsi's model of affect falls into the same category as these approaches, and uses the following six dimensions, based on Dörner's Psi theory (see figure 5):

- *Valence:* the positive or negative reinforcement signals resulting from the satisfaction or frustration of demands.
- *Arousal:* proportional to the urgency and importance of the currently active demands, it is the equivalent of the *unspecific sympathicus syndrome* in humans, and increases goal directedness. In biological systems, arousal also controls the allocation of physiological resources, for instance, it diverts oxygen to increase muscular responses at the cost of digestive processes, with implications for the proprioceptive component of emotions (heart rate, stomach pains, tension, sweating).
- *Resolution level:* controls the speed and accuracy of perception, memory access and planning by adjusting the width and depth of activation spreading in the agent's representations. A high resolution level results in slow processing, but a deep and detailed perceptual/memory exploration, while a lower resolution level results in fast processing and a penalty on accuracy and detail. The resolution level is driven

up by a high urgency of the leading motive (resulting more focused cognition), and decreased by the activity of the demands in general. Also, the resolution level is inversely related to the arousal: a high arousal requires faster processing, while a low arousal allows more attention to detail.

- *Selection threshold:* if the currently dominant demand is very urgent, it is important to avoid goal oscillations, i.e. to keep the current behavior directed upon the satisfaction of this demand. The stability of the behavior is controlled by adding a 'bonus' weight to the current goal, so it becomes harder for competing demands to become dominant. The selection threshold amounts to an adaptive "stubbornness". Additionally, a low estimate of the agent's ability to reach its goals (competence) will reduce the selection threshold and thus enable greater flexibility.
- *Goal directedness*: this parameter balances explorative/deliberative vs. executive strategies. A high urgency of active motives and a high arousal increases goal directedness, while a high level of uncertainty about the environment decreases it.
- *Securing rate:* controls the rate of background checks of the agent, and thus balances between attentive perception and other cognitive processing. A high perceived uncertainty of the environment involves a high securing rate, while a high task-specific competence will lower it. Also, a high importance of the leading motive leads to a reduction in background checks, and thus in a lower securing rate, too.

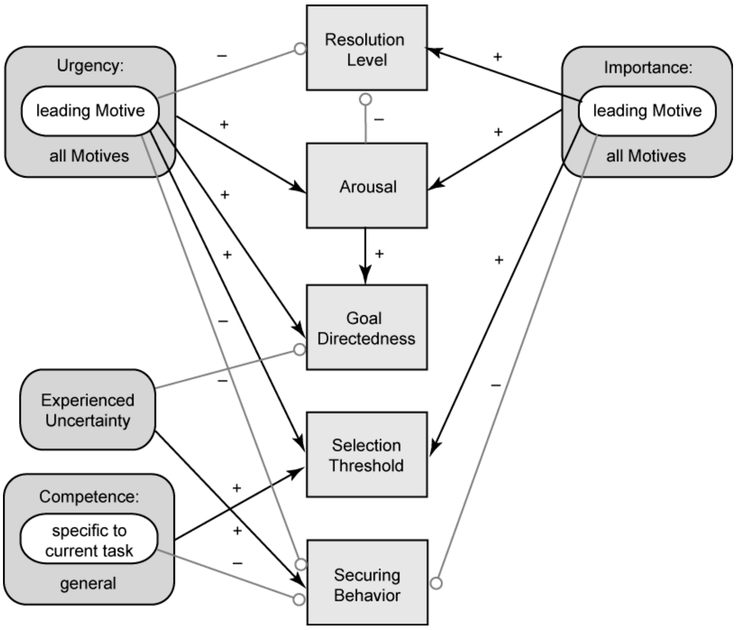

Figure 4: Affective dimensions according to the Psi theory (adopted from Hille, 1998, see also Bach, 2009, p. 149)

Obviously, these dimensions are not orthogonal, but dependent on each other to a large degree. The exact mode of dependence (i.e., the quantitative influence between them) is likely a source of some differences in individual personality.

The modulators suggested by the Psi theory, and used in the MicroPsi framework, can be interpreted as a super-set of the Wundt, Traxel/Heide and Mehrabian/Russel models: valence and arousal are equivalent, the submission/dominance dimension is represented as general competence (see Mehrabian 1980 for an analysis of the conceptual similarity between dominance and competence), while tension/stress may be decomposed into securing behavior (activation due to high uncertainty) and the consequences of arousal due to motivational urgency.

We can use the described dimensions to characterize affective states. For instance, anger amounts to a high arousal, low resolution level, strong goal dominance (a high selection threshold), strong goal directedness and few background checks. Sadness is characterized by a low arousal, a high resolution level, few background-checks and low goal-directedness. The modulators also account for much of the hedonic aspect of affective states, i.e., the specific way emotion is reflected via proprioception. For example, the high arousal of anger will lead to heightened muscular tension and a down-regulation of digestion, which adds to the specific way anger feels to an individual. Modulators can also describe subtle emotional differences, like the one between enthusiastic *joy* and quiet *bliss*: both are characterized by strong positive valence, but bliss is accompanied by a low arousal and high resolution level.

However, while the modulator model on its own cannot address higher level emotions, for instance anger about someone, grief over a loss, pride about an achievement, or disappointment over an event: affective states that are directed upon an object will require a motivational system that supplies the specifics of that directedness.

The motivational system: generating relevance from demands

As mentioned in section 3, the motivational system can be characterized by a (pre-defined) set of *demands* of the system, which are represented to the cognitive architectures as *urges*: A *drive* is a demand, represented by an urge signal. Changes in these signals determine *valences*: a change of a demand towards its target value creates a positive reinforcement (*pleasure* signal), while a negative change away from the target results in a negative reinforcement (*displeasure* signal). These signals can be used to create associations between the urges and situations that satisfy them (*goals*) or frustrate them (*aversive situations*). The association between urges and situations is represented as a weighted link (figure 5) that allows the retrieval of situations by spreading activation: If an urge becomes active, all associated situations will be pre-activated, i.e., the memory of the agent will be primed for the retrieval of those situations that afford the satisfaction of the urge. (For a more detailed explanation of the learning and retrieval mechanism, see Bach, 2009.)

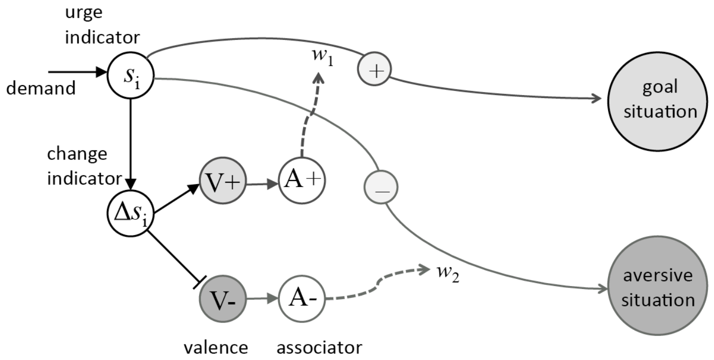

Figure 5: Associating urges with goal situations

In an open environment, with an unlimited supply of unknown situations and events, it is impossible to specify a complete set of goals a priori. Using a demand-based system solves this problem, by letting the agent identify and establish its goals autonomously, based on its systemic needs. Just as the cognitive modulators specify affective dimensions, the demands determine motivational dimensions.

Note that a definition based on a pre-defined *utility function* is not helpful in the context of an architecture of cognition. Utilities quantify the relationship between goals and demands, but in reality, this relationship is fraught with contingencies and irregularities that can only be captured by looking at the mechanisms of goal selection in detail. For example, a cognitive mechanism could bias an agent in such a way that the color of a room might have an influence on the choice of food. A utility function that reflects this would have to encompass this mechanism, and in the worst case the whole cognitive architecture, which renders the concept of utilities meaningless. (Utilities have their place in domain-specific, externalist agent abstractions, for instance in economic simulation.)

To enable social and cognitive behaviors that go beyond sub-goals of the pursuit of food and pain-avoidance, an agent needs genuine social and cognitive demands. Since it is not clear how to derive or acquire them indirectly (as proposed by Sun, 2005), or identify them in an abstract sense (see Maslow et al., 1987), the Psi theory specifies them explicitly, and on the same level as physiological drives.

In accordance with the Psi theory, MicroPsi uses three groups of demands: physiological, social and cognitive.

The *physiological demands* (food, water, physical integrity/pain avoidance etc.) become active whenever the autonomous regulation of physiological parameters fails and provide for the basic survival. Here, survival itself is seen as an abstract concept and not a demand itself.

Social demands consist in a need for affiliation with others, and are mediated by social signals (‘legitimacy signals’), such as displays of affection, acceptance, rejection or reproach. The affiliation mechanism allows to structure social interaction beyond rational utility: purely social rewards are often sufficient to motivate an agent for cooperative behavior, without incurring the need to supply a material gratification and thereby affect the fitness of the group, or to discourage anti-social behavior without decreasing the agent's material fitness by doling out punishment.

A second social demand is called 'internal legitimacy': it corresponds to internal social signals that are related to the conformance to internalized social norms ('honor'). Obviously, the list of social demands is incomplete; for instance, it lacks sexual needs (libido). MicroPsi's implementations, both for simulations and for robots, did not offer any opportunities to address these.

The group of *cognitive demands* spans needs for competence, a need for uncertainty reduction, and needs for aesthetics. Let us look at the cognitive demands in more detail.

Competence is either *epistemic* (related to skills): it provides an estimate on the agent's ability to cope with any *specific* task, by delivering a reward on its successful completion, and a penalty on failures. Thus, skill-acquisition can become a goal on its own. Furthermore, competence may be *general*, i.e. related to the overall ability of the agent to cope with the environment. General competence delivers a heuristics on the amount of risk an agent should take, and is measured as a floating average over successes and failures of the agent's past actions. Finally, competence might be *effect related*: the ability to produce a large visible change in the environment produces a reward signal on its own.

While the three kinds of competence are evaluated differently, they share a reward system and are to some extent interchangeable: if an agent perceives a very low subjective level of general competence (because many of its actions fail, or because it anticipates failure of a difficult task, such as writing a research paper), it may compensate by producing a large effect (for instance, by destroying something), or by resorting to the execution of a skill where it possesses a high epistemic competence (such as cleaning the dishes, or surfing the internet: a pattern of behavior known as *procrastination*).

Uncertainty reduction is aimed at discovering the outcomes of actions, and exploring the structure of objects and situations. Sometimes, this is addressed by a drive for *novelty*. I prefer the concept of uncertainty reduction, because obviously, many people avoid novelty in areas where they are not competent. Novelty seeking takes place precisely in those cases where uncertainty reduction is anticipated. Also note that the goal is not a situation where uncertainty permanently disappears: the reward is given for the reduction of uncertainty, not for its absence. Thus, agents will usually aim for situations with exploratory potential, but within their competence of successful exploration.

Uncertainty reduction is satisfied by '*certainty events*': the complete identification of an object, scene or frame; by fulfilled expectations (even negative ones), and by a long and non-branching expectation horizon. Conversely, uncertainty reduction is frustrated whenever the agent encounters unknown objects or events, discovers elements without a known connection to behavior (the agent has no knowledge what to do with them), when there is difficulty to perceive or resolve the current situation at all, expectations have been breached, or the current expectation horizon is too short or branches too much, so that predictions of future events are difficult.

Uncertainty signals are weighted with the motivational relevance of their object. Generally, a high uncertainty will give rise to explorative behaviors, unless the agent has a low epistemic competence for exploration.

Aesthetics is a demand that directs the agent at seeking order, i.e. better representations (*abstract aesthetics*), or seeking out particular stimuli, based on evolutionary preferences, such as certain body schemas or landscapes (*stimulus oriented aesthetics*).

The selection of motives

All goals of a MicroPsi agent are defined by a *consumptive action*, i.e. the satisfaction of one (or more) of the aforementioned needs. Usually, consumptive actions are embedded into environmental situations that afford them, but they are not restricted to physiological effects. For the cognitive system of the agent, the displeasure signal incurred from a negative social signal is just as real and action-relevant as the displeasure signal received from a physical injury.

Each demand is characterized by several parameters (figure 6):

- The *target value* v_d of the demand *d*
- The *deviation* $| v_d - c_d |$ from that value, represented by an urge indicator $urge_d$,
- The *weight* of the demand (its relative importance, compared to other demands with the same urgency) w_d,
- The *gain* (the satisfaction derived from a positive stimulus or consumption) g_d,
- The *loss* (the penalty incurred from a negative stimulus or a frustration) l_d,
- The *decay* (the autonomous increase of the deviation from the target value over time) f_d.

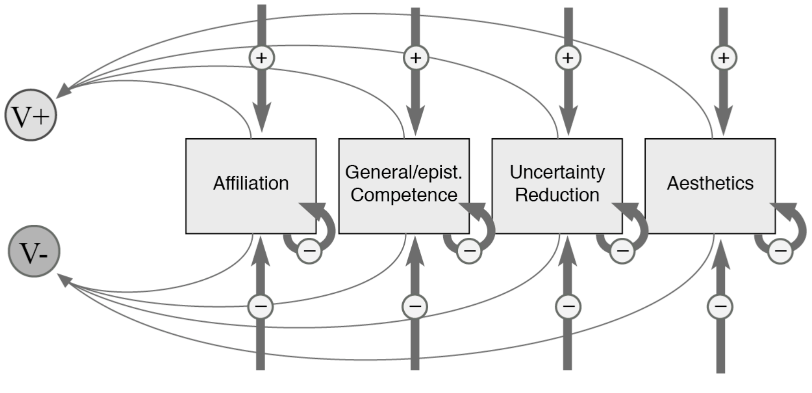

Figure 6: Motivational dynamics: Gain, loss and decay of drives (note that the interrelations between competence and satisfaction of other drives not shown)

Even if no gain or loss is incurred, the decay ensures that the motivational parameters change relentlessly, and the agent is requiring to constantly replenish the demands. At each point in time, the active demands are weighted against each other. *Motives* are active demands, combined with the associated goal situations that afford their satisfaction. The agent will attempt to satisfy active demands opportunistically, and failing that, it will attempt to retrieve or construct a plan based on its memory content.

Following a plan requires a commitment, which is represented by elevating a motive to an *intention*.

An intention in the context of our cognitive architecture does not necessarily carry the same connotations as the notion of *intentionality* in the philosophy of mind. Here, we use the term in much the same sense as in the context of *belief-desire-intention* models. In that terminology, a *motive* is a *desire*, and '*intention'* is a leading motive; it simply refers to the set of representations that initiates, controls and structures the execution of an action. Note that intentions may form *intention hierarchies*, i.e. to reach a goal it might be necessary to establish sub-goals and realize them one after the other. Therefore, an intention can be specified by a goal state, an execution state, an intention history (the protocol of operations that took place in its context), a plan, the urge associated with the goal state (which delivers the relevance), the estimated specific competency to fulfill the intention (which is related to the probability of reaching the goal) and the time horizon during which the intention must be realized (figure 7).

Intentions will be selected based on their expected success probability, multiplied with their importance. This means that it is possible that even though a motive is very active, it might never become an intention, if the environment does not allow its satisfaction (i.e., the probability of satisfying it is near zero). As a result, the agent incurs relentless displeasure signals from the frustration of the associated demand, but won't be able to address these.

The importance of a motive is simply given by the weighted strength of the associated urge. The probability of succeeding is given by the task-specific competence, calculated for a particular sequence of actions (plan). The acquisition of skills for reaching goals is a cognitive demand in its own right, but often, there might be no valid estimate available for reaching a particular goal situation, especially if the agent attempts to perform a sequence of actions for the first time. Here, the agent uses its general success rate, which is given as the *general competence*, as a heuristics. Thus, the chance of reaching a particular goal can be approximated as the sum of the general competence and the epistemic competence for that goal, and the *motive strength* to satisfy a need *d* is calculated as $w_d \cdot urge_d \cdot (generalCompetence + competence_d)$, i.e. the product of the relative strength of the urge and the combined competence.

As mentioned above, the selection of motives is modulated by the agent's selection threshold. The selection threshold is an adaptive value that is added as a 'bonus' to the activity of the currently selected intention, making it harder for competing motives to take the lead. Without a selection threshold, the agent would be prone to oscillation between active motives without following a plan long enough to satisfy any one of them.

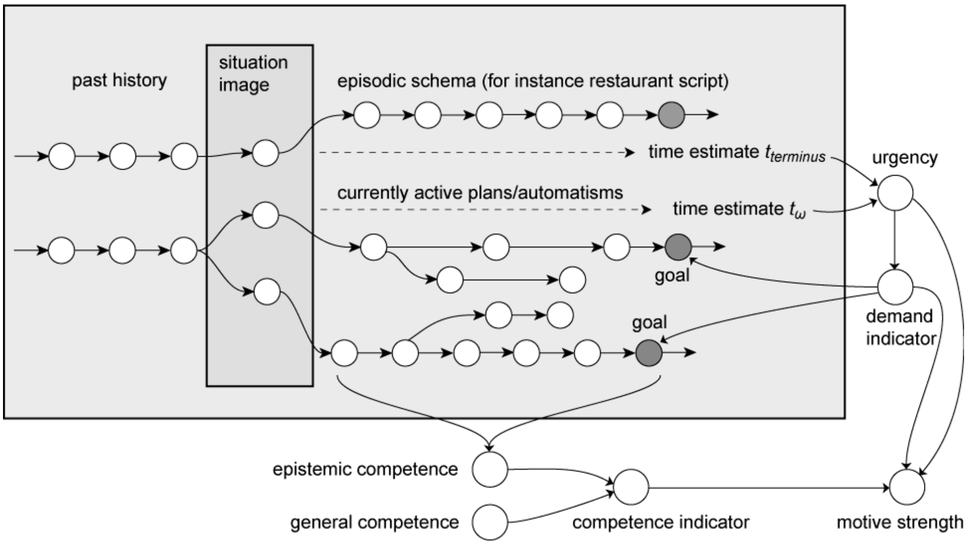
past history
situation image
episodic schema (for instance restaurant script)
time estimate $t_{terminus}$
currently active plans/automatisms
time estimate t_{ω}
urgency
goal
goal
demand indicator
epistemic competence
general competence
competence indicator
motive strength

Figure 7: The structure of a motive (Bach, 2009, p. 131)

Applying the MicroPsi framework for modeling personality traits

The motivational architecture described in the previous sections is a *qualitative model*; while it specifies the structural relationships of its components, it does not quantify them. A commitment to specific values for the individual weight parameters allows for the modeling of agents with specific individual properties. For example, by varying the weights w_d for set of demands, it is possible to produce behavior that is either more explorative, more resource oriented, or more cooperative. Evolutionary simulations suggest that there is no single optimal setting, but that different environments favor different combinations for those weights (Dörner et al., 2006).

The systematic exploration of the influence of different settings not just for the demand weights, but for the motivational dynamics themselves is the next logical step. Thus, the motivational traits of agents can be defined as a set of physiological, social and cognitive demands *D*, each of them annotated by a tuple (w_d, g_d, l_d, f_d), describing the weight, gain, loss and decay of the respective demand.

Using these parameters, it is possible to create agent models that conform to the Five Factor Model (*FFM*, or *"Big Five"*) established in personality psychology. The FFM suggests five dimensions of personality traits, which together can be used to characterize emotional/motivational dispositions of an individual (Digman, 1990; Goldberg, 1993). These trait dimensions are usually called:

- *Openness:* This describes the interest a subject takes in new situations, ideas and stimuli. Openness is associated with intellectual curiosity, appreciation of art, and non-conservatism
- *Conscientousness:* This characterizes how organized/rigid a subject tends to be. Conscientous individuals tend to spend more time planning, attend carefully to details and attempt to follow plans and rules rigorously.
- *Extraversion:* This relates to the interest individuals take in interpersonal interaction, their surgency and expressiveness.
- *Agreeableness:* Individuals that are highly agreeable tend to avoid conflicts, are friendly and seek positive social interaction.
- *Neuroticism:* This amounts to emotional instability. Subjects with a high degree of neuroticism tend to experience negative emotions more strongly, are prone to anxiety and mood switches.

Each aforementioned property marks just one end of the respective trait dimension, of course. For instance, the extraversion axis ranges from extreme introversion to extreme extraversion, the agreeableness axis from disagreeable, conflict seeking behavior to pronounced conflict avoidance, and so on. An individual would be characterized by five values, each one quantifying a particular value on one of the five scales, to specify the expression or respective absence/inversion of the trait.

Modeling configurations of personality traits by choosing appropriate settings for the tuples (w_d, g_d, l_d, f_d) is straightforward. Since all of them are related to social and cognitive pre-dispositions, it is sufficient to look at the demands for *affiliation*, *competence*, *certainty* (= uncertainty reduction) and *aesthetics*.

For instance, a high degree of neuroticism can be expressed by choosing particularly high values for the loss and decay of *competence* and *certainty* (and possibly the other demands, too). In other words, the agent needs to replenish its competence and certainty very often, and it will react disproportionally to failures of doing so, and to frustrations of these demands. The continuous decay of *certainty* makes the agent prone to episodes of anxiety (figure 8). Conversely, an agent with the opposite settings, i.e., very low decays and losses on *competence* and *certainty* will not take a big hit on failure, and won't need to seek out new competence and certainty rewards as often. Thus, it will display a greater degree of emotional stability and complacency.

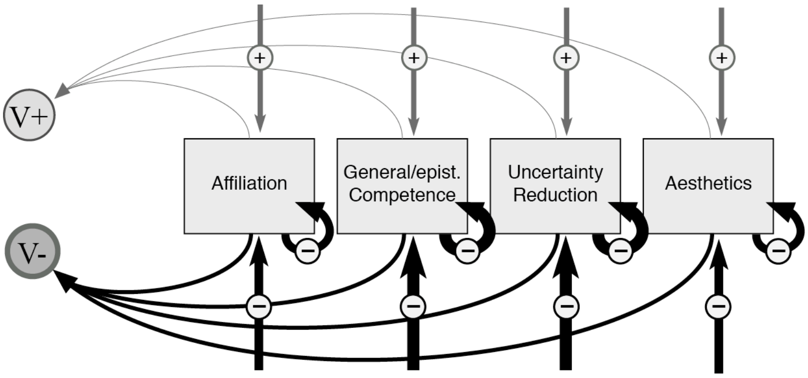

Figure 8: Dynamics of neuroticism

Compare this with the dynamics of a highly *open* agent (figure 9): Here, we have a high decay on *competence* and *certainty*, too, so the agent is forced to seek out a lot of competence and exploration rewards. On the other hand, it receives a high gain on satisfying its cognitive (and possibly social) demands. Thus, it will receive positive frequent and strong positive reinforcements of its explorative and competence building behaviors, resulting in a high tendency to seek out new situations and stimuli.

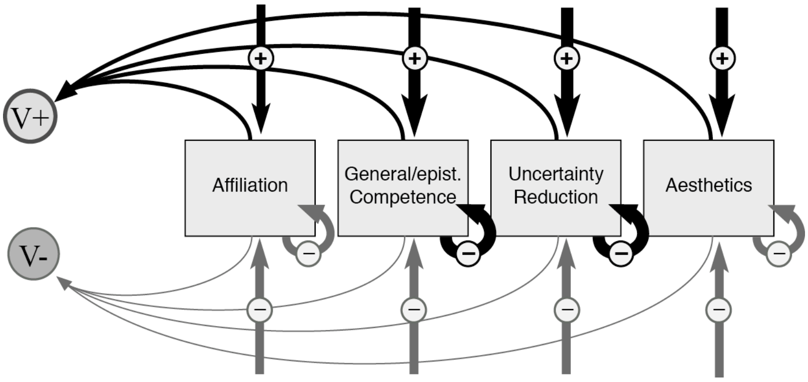

Figure 9: Dynamics of openness

Our model determines *conscientousness* with a strong loss factor of *competence* and *certainty*, combined with a weak gain of *competence*/*certainty*. This means that the reward for exploration and skill acquisition is low, compared from the loss incurred by risking them. A high decay on *competence*, but low decay on the other drives can additionally result in a low interest in seeking out new social, aesthetic or exploratory challenges, while focusing on a high accuracy in the execution of plans and skills (figure 10). Additionally, we may model rigidity with a higher value for the *selection threshold* modulator. Calm, conscientious agents may also receive a low increase of *arousal*/reduction of *resolution level* due to demand activity.

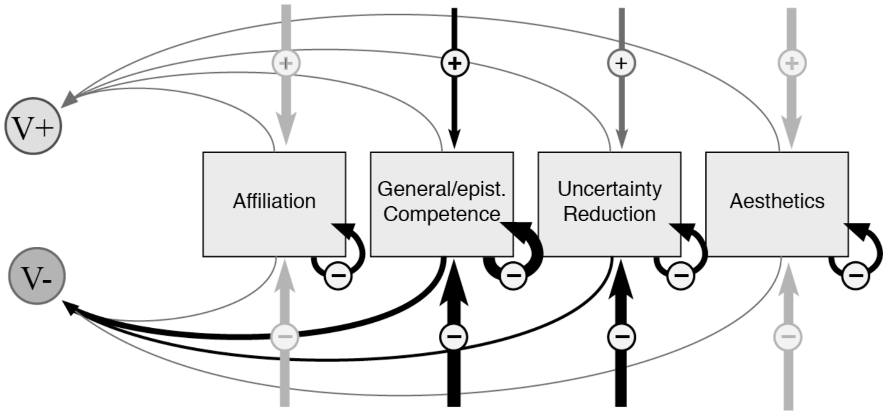

Figure 10: Dynamics of conscientiousness

Extraversion is produced by a high decay of the *affiliation* demand, which therefore requires constant social interaction to be replenished (figure 11). Strong gains on *affiliation* and *competence*, as opposed to weak losses on these drives result in a strong reinforcements due to social and competence successes, but only little aversion due to failures.

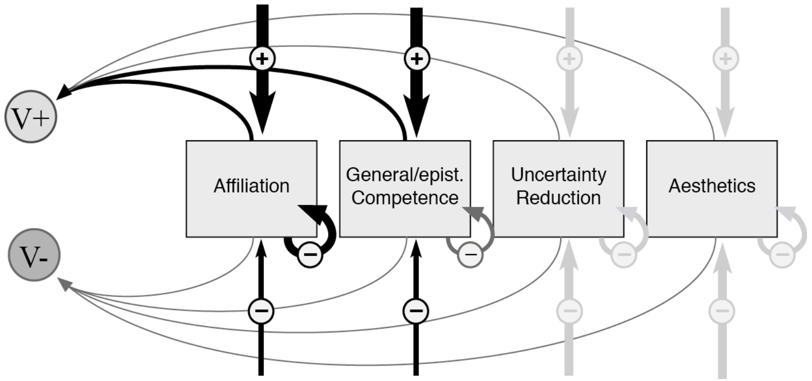

Figure 11: Dynamics of extraversion

Agreeable agents are somewhat similar to extroverts due to a high decay on *affiliation* (and possibly *competence*), so they need to seek out social situations often. Unlike extroverts, they receive strong *affiliation* losses due to negative social signals, and gain little *competence* (figure 12). Thus, they are likely to avoid arguments: they have little positive rewards to gain from them, but incur strong negative reinforcements if they do not succeed socially.

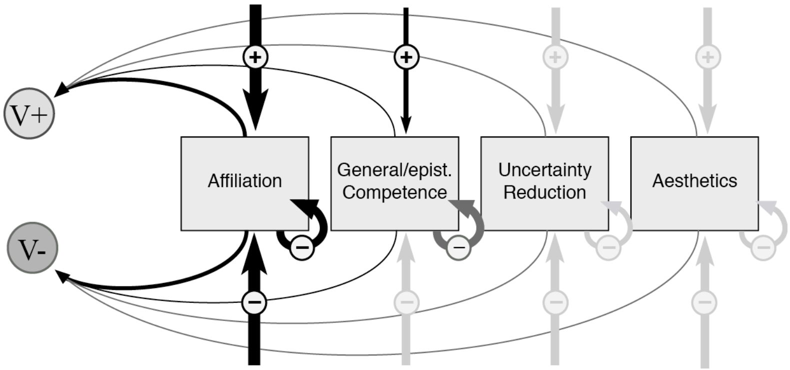

Figure 12: Dynamics of agreeableness

While the MicroPsi framework is well suited to capture the desired traits, it involves vastly many more free variables, which is very unfortunate for a rigid experimental evaluation. Even if we only look at the subset discussed in this section, we require four parameters to specify the dynamics of four demands, so that 16 free variables are used to capture the five dimensions of the FFM. Arbitrarily fixing these values does not present a good strategy out of this dilemma, because it introduces additional assumptions without a theoretical or experimental justification. Instead, such a reduction might needlessly reduce the expressivity of the framework, especially since the top-level description of the FFM is not a complete description, but only a rough approximation of human personality traits. Take for example *shyness*: we might be tempted to subsume shy behavior under the headline of introversion, because both shy and introverted individuals tend to receive strong negative rewards from failed social interaction, and therefore tend to shun it. But not all shy people are introverts, which do not gain positive rewards from successful social interaction. Instead, they tend to have a low self-attributed competence for handling social situations, which is strongly reinforced by failures (see figure 13). Thus, a shy individual could also be a closeted extrovert, secretly craving positive social signals.

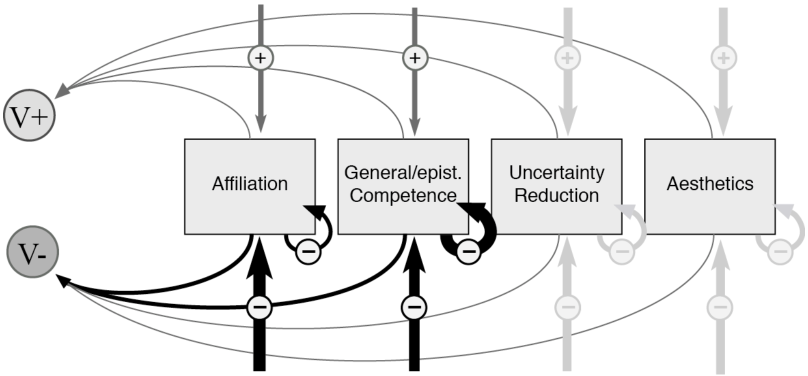

Figure 13: Dynamics of shyness

Because of its qualitative nature, the MicroPsi framework is not so much a theory that could be tested in a series of psychological experiments. Like the BDI framework, it offers a parsimonious way to capture its problem domain *conceptually*, and it does this with a sufficient degree of resolution to implement it as a computational model of motivation, affect and emergent emotional states and processes. In other words: MicroPsi is an attempt at a conceptually minimal model of the motivational and emotional aspects of cognition.

MicroPsi agents have been implemented in virtual environments and used for the control of robots. However, our group aims at anchoring the model in experimental research, especially in the domain of psychometrics of personality traits. Since the performance of humans and computational models with respect to established tests for personality traits is not comparable, we are using our model to design a series of problem solving scenarios that correlate personality properties with the performance of subjects (Greiff & Funke, 2009). As a result, we hope to provide a direct application of the model for psychometric purposes. Furthermore, well-defined problem solving scenarios present an opportunity to compare the performance of human subjects directly with that of computational agents and will thereby force us to revise the motivational and emotional framework of the cognitive architecture presented here.

Acknowledgments

The work presented here is supported by a postdoctoral fellowship grant of the Berlin School of Mind and Brain, Humboldt University of Berlin, and benefits from the contribution of several collaborators and colleagues. I would especially like to thank Dietrich Dörner, Joachim Funke and Andreas Fischer.

References

Bach, J. (2003). *The MicroPsi Agent Architecture*. Proceedings of ICCM-5, International Conference on Cognitive Modeling, Bamberg, Germany, 15-20

Bach, J. (2007). *Motivated, Emotional Agents in the MicroPsi Framework*. 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.

Bach, J. (2011). *A Motivational System for Cognitive AI*. In Schmidhuber, J., Thorisson, K. R., & Looks, M. (eds.): Proceedings of Fourth Conference on Artificial General Intelligence, Mountain View, CA. 232-242

Becker-Asano, C. (2008). *WASABI: Affect Simulation for Agents with Beliebable Interactivity.* PhD Thesis, Faculty of Technology, University of Bielefeld

Bratman, M. (1987). *Intentions, Plans and Practical Reason.* Harvard University Press

Castelfranchi, C., & Miceli, M. (2009). *The cognitive-motivational compound of emotional experience.* Emotion Review, 1, 223-231

Cole, P. M., Michel, M. K., & O'Donnel Teti, L. (1994). *The Development of Emotion Regulation and Dysregulation: A Clinical Perspective.* Monographs of the Society for Research in Child Development, Vol. 59: 2/3 Blackwell Publishing: 73-100

Dias, J., & Paiva, A. (2005). *Feeling and Reasoning: A Computational Model for Emotional Characters.* In: Bento, C., Cardoso, A., Dias, G. (eds.) EPIA 2005. LNCS 3808, Springer: 127-140

Diener, E. (1999). *Special section: The structure of emotion.* Journal of Personality and Social Psychology, 76, 803-867

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: Rowohlt

Dörner, D., Bartl, C., Detje, F., Gerdes, J., & Halcour, D. (2002). *Die Mechanik des Seelenwagens. Handlungsregulation.* Verlag Hans Huber, Bern

Dörner, D., Gerdes, J., Mayer, M., & Misra, S. (2006). *A Simulation of Cognitive and Emotional Effects of Overcrowding.* In: Fum, D., de Missier, F. & Stocco, A. (eds.): Proceedings of the Seventh International Conference on Cognitive Modeling (ICCM 2006). Trieste: Editione Goliardiche: 92 – 99

Ekman, P., & Friesen, W. (1971). *Constants across cultures in the face and emotion.* In: Journal of Personality and Social Psychology 17(2): 124-29

Ellsworth, P. C., & Scherer, K. R. (2003). *Appraisal processes in emotion.* In Davidson, R. J., Goldsmith, H. H., & Scherer, K. R. (eds.) Handbook of the affective sciences. New York, Oxford University Press

Ertel, S. (1965). *EED - Ertel-Eindrucksdifferential (PSYNDEX Tests Review).* Zeitschrift für experimentelle und Angewandte Psychologie, 12, 22-58

Frijda, N. H. (1986). *The emotions.* Cambridge, U.K., Cambridge University Press

Frijda, N. (1987). *Emotion, cognitive structure, and action tendency.* Cognition and Emotion, 1, 115-143.

Goldberg, L. R. (1993). *The structure of phenotypic personality traits.* American Psychologist 48 (1): 26–34

Gratch, J., & Marsella, S. (2004). *A framework for modeling emotion.* Journal of Cognitive Systems Research, Volume 5, Issue 4, 2004, p. 269-306

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

Hudlicka, E., & Fellous, J.-M. (1996). *Review of computational models of emotion* (Technical Report No. 9612). Psychometrix. Arlington, MA

Izard, C. E. (1994): Innate and universal facial expressions: Evidence from developmental and cross-cultural research. Psychological Bulletin, 115, 288-299

Lazarus, R. (1991) *Emotion and Adaptation,* NY, Oxford University Pres

Lisetti, C., & Gmytrasiewicz, P. (2002). *Can a rational agent afford to be affectless? A formal approach.* Applied Artificial Intelligence, 16, 577-609

Marsella, S., Gratch, J., & Petta, P. (2011). *Computational Models of Emotion.* In Scherer, K. R., Bänziger, T., & Roesch, E. (eds.). *A blueprint for an affectively competent agent: Cross-fertilization between Emotion Psychology*, Affective Neuroscience, and Affective Computing. Oxford University Press

Maslow, A., Frager, R., & Fadiman, J. (1987). *Motivation and Personality.* (3rd edition) Boston: Addison-Wesley

Mehrabian, Albert (1980). *Basic dimensions for a general psychological theory*. Oelgeschlager, Gunn & Hain Publishers: 39–53

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge, U.K., Cambridge University Press.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press

Peña, L., Peña, J.-M., & Ossowski, S. (2012). *Representing Emotion and Mood States for Virtual Agents.* In Klügl, F., & Ossowski, S. (eds.): MATES 2011, LNAI 6973, Springer: 181-188

Plutchik, R. (1994). *The Psychology and Biology of Emotion.* New York: Harper Collins

Reisenzein, R. (2001). *Appraisal processes conceptualized from a schema-theoretic perspective: Contributions to a process analysis of emotions.* In K. R. Scherer, A. Schorr & T. Johnstone (eds.), Appraisal processes in emotion: Theory, methods, research, Oxford: Oxford University Press. 187-201

Roseman, I. J. (1991). *Appraisal determinants of discrete emotions.* In: Cognition and Emotion, 3, 161-200

Russel, J. A. (1995). Facial expressions of emotion. What lies beyond minimal universality. Psychological Bulletin, 118, 379-391

Schlosberg, H. S. (1954). *Three dimensions of emotion.* Psychological Review 1954, 61, 81-8

Sloman, A. (1981). *Why robots will have emotions.* Proceedings IJCAI.

Sloman, A., Chrisley, R., & Scheutz, M. (2005): *The Architectural Basis of Affective States and Processes*. In Fellous, J.-M., & Arbib, M. A. (eds): Who needs emotions? The Brain meets the robot, Oxford University Press, p. 203-244

Sun, R. (2004): *Desiderata for Cognitive Architectures.* Philosophical Psychology, 17(3), 341-373

Sun, R. (2005): *Cognition and Multi-Agent Interaction.* Cambridge University Press, 79-103

Woodworth, R. S. (1938). *Experimental Psychology*. H. Holt and Company

Wundt, W. (1910). *Gefühlselemente des Seelenlebens.* In: Grundzüge der physiologischen Psychologie II. Leipzig: Engelmann D. R. Bates, *Phys. Rev.*, 492 (1950)