My Agents Love to Conform: Emotions, Norms, and Social Control in Natural and Artificial Societies

Christian von Scheve
ZiF, University of Bielefeld
Wellenberg 1, D-33615 Bielefeld, Germany
xscheve@informatik.uni-hamburg.de

Daniel Moldt
University of Hamburg, Computer Science Department
Vogt-Koelln-Str. 30, D-22527 Hamburg, Germany
moldt@informatik.uni-hamburg.de

Julia Fix
University of Hamburg, Computer Science Department
Vogt-Koelln-Str. 30, D-22527 Hamburg, Germany
julia.fix@gmx.de

Rolf von Lüde
University of Hamburg, Institute of Sociology
Allende-Platz 1, D-20146 Hamburg, Germany
luede@uni-hamburg.de

Abstract

This contribution investigates the function of emotion in relation to social norms, both in natural and artificial societies. First, the authors briefly illustrate that norms as socially shared mental objects play a crucial role in the dynamics of social structures and social order, in natural societies as well as in artificial systems. Second, the authors address the question how norms are enforced and thereby maintained throughout a social system. In this respect, it is shown that emotions play a crucial role by providing means for intrinsic gratification and sanctioning. The authors consider emotion related sanctions as a cost equivalent to and in many situations perhaps even more efficient than, e.g., resource-driven penalties. Consequently, agents' anticipation of negative emotional outcomes as a consequence of deviant behaviour is supposed to exert social control. Third, the authors outline the possibilities of an application to the socionic multi-agent architecture SONAR

1 Introduction

For some time now apprehensions from the general public as well as from the scientific community have been issued concerning the controllability of artificial intelligence systems, in particular distributed systems based on intelligent autonomous agents. Agent systems are feared to run out of control in such a way that autonomously generated (although probably temporary) goals pursued by a system might contradict (implicit) high-level goals of the designer or user, respectively. Unfortunately, the dilemma arising out of these possible goal conflicts affects some of the core strengths of artificial agent systems: autonomy, flexibility, and discretion. Therefore, means have to be developed that on the one hand ensure the autonomy of the systems in question, and on the other hand avoid conflicts with implicit human high-level goals.

One solution to this problem is the implementation of a system of socially shared norms which is not coerced by the designer, but instead emerges from the mutual interactions of the agents (with actors and/or users). To realise such an approach, it would be beneficial, if not mandatory, to have profound knowledge of and adapt to the computational context the mechanisms of norm emergence, prevalence, and compliance in human social systems (e.g., Dignum et al. 2000; Saam/Harrer 1999).

We argue that *emotions* constitute such a mechanism and that they should therefore be taken into account in the design of agents and especially multi-agent systems (MAS). However, the concept of agents is inspired by and to a large extent also relies on findings from (cognitive) psychology and hence on this discipline's conceptualization of intelligent behaviour, which still is fundamentally based on cognition. Belief-Desire-Intention (BDI) architectures can well be considered an epitome of this perspective on intelligent information processing.

Notwithstanding this, the interrelation of emotion and cognition and the role of emotion in overall intelligent behaviour have been long debated in psychology and have likewise promoted the idea that artificial intelligence (AI) systems could be improved by taking into account mechanisms which are functionally equivalent to emotion in biological systems (Simon 1967; Sloman/Croucher 1981).

At least since Marvin Minsky's programmatic and frequently cited statement that „the question is not whether intelligent machines can have emotions, but whether machines can be intelligent without any emotions" (Minsky 1986: 163), efforts have been increased within the AI community to develop "emotional agents", i.e. software agents capable of utilising mechanisms which are functionally equivalent to emotions in human and non-human animals (for extensive overviews cf. Cañamero (1998), Hatano and associates (2000), as well as Trappl and colleagues (2003)). Unfortunately, until now, the function of emotions in larger societal structures has not been investigated thoroughly, despite some efforts in the area of distributed systems research (e.g., Elliot (1993), Aubé/Senteni (1996), Gmytrasiewicz/Lisetti (2000), and Fix (2004); see also von Scheve/Moldt (2004)).

In view of the fact that distributed (AI) systems are of increasing importance in many areas of application, e.g., electronic marketplaces, automated negotiations, planning and scheduling systems, business process and workflow management, coordination of large-scale open systems, and simulations, it seems reasonable to further investigate the function of emotion in large scale social systems, natural and artificial ones. Pioneering research in the computational study of social norms and emotion has been conducted by Alexander Staller and Paolo Petta (2001); however, this contribution focuses slightly different goals in that it emphasises social structural dynamics.

The article is structured as follows: In the next section we first illustrate the social functions of emotion, both in view of an agent's internal functioning and in view of social interactions. In section three we argue for a model of social control that is fundamentally based on two specific functions of emotion in relation to social norms: the triggering of action incentives (action readiness) and the control of social action. In this model, the emotional commitment to norms in particular ensures actor's compliance with these norms. In the fourth section we then outline how these findings might be applied to the Sonar multi-agent architecture.

2 Social Functions of Emotion

This section examines the functions of emotion in social interactions. The social functions of emotions can conceptually be distinguished from their intraindividual (Levenson 1999), phylogenetic (Turner 2000; Cosmides/Tooby 2000) or ontogenetic functions (Abe/Izard 1999; Holodynski/Friedlmeier 2005).

In order to perform a functional analysis of emotion, we first sketch our definition of emotion and our understanding of functional analysis. Thus, we define emotion as „functional, organised responses to environmental demands that prepare and motivate the person to cope with the adaptational implications of those demands" (Smith/Pope 1992: 36), whereas environmental demands in principal can be both, physical and social. According to this definition, a central function of emotion is the adaptational and beneficial regulation of an agent's behaviour in relation to its environment (Keltner/Gross 1999: 468).

Averill (1992), for example, locates these social functions on three different levels (biological, psychological, and social), Keltner and Haidt (1999) on four levels (individual, dyadic, group, and cultural), Gerhards (1988) likewise on four levels (organism, personality, social structure, and culture) and von Scheve and Moldt (2004) on three levels of analysis (micro, meso, macro). These partitions differ in principle only conceptually and in their ratio of abstraction and/or reduction (cf. also Turner 2002: 30-41).

On the level of an individual agent, emotion performs above all two functions: on the one hand, emotion informs an agent about those events in the social environment that often require immediate, reactive and adaptive behaviour (Schwarz 1990; Clore et al. 1994). For example, annoyance informs about the felt fairness of an action; love informs about degrees of affection and commitment; shame and embarrassment inform about the conformity of an action (cf. Keltner/Haidt 1999). On the other hand, emotions prepare an agent to react adequately upon requirements arising from social situations, e.g. through physiological changes (Clore 1994; Dimberg 1997; Cacioppo et al. 2000). An analogous view is put forward by Oatley and Jenkins (1996: 252), who locate the intraindividual functions of emotion in action readiness and in the structuring of the cognitive system into adequate operational modes.

On the interindividual level, i.e. the expressive and communicative level, there are above all three social functions of emotion that relate in particular to social norms (see Keltner/Haidt 1999):

First of all, emotion expressions allow the attribution of most interactional contingencies, including emotional state, appraisals, intentions, and corresponding interpretations of the situation.

Secondly, emotion expressions may (unconsciously) evoke complementary or reciprocal reactions in context-bound observing actors and therewith contribute to improved bilateral interpretations of a situation. This in turn is a prerequisite for cooperation and the coordination of action.

Thirdly, emotion expressions promote or obstruct specific courses of action and interaction for interacting individuals by exhibiting either motivating or sanctioning clues.

In view of the social environment, within which the agent-environment contingencies and reciprocities emerge, Keltner and Haidt (1999: 507) argue: „Functional explanations refer to the history of some object (e.g. behaviour or trait), as well as the regular consequences that benefit the system in which the object or trait is contained". We insofar attach our argument to this viewpoint as we focus the *regular consequences* for the system that contains an agent as well as that agent's actions which are in part guided by norm related emotions.

The social functions of emotion in larger social units can be seen at their contributions to identifying social groups and group members (Durkheim 1994), at the ascription of status and power resources (Kemper 1978), at the construction and maintenance of solidarity and cohesion (Lawler et al. 2000), and at the internalisation and retention of social norms, power structures, moral ideas, and ideology concepts (Elster 1999; Hochschild 1979/1983).

Having briefly clarified our position regarding the social functional analysis of emotion, we move towards a more general social scientific analysis of social functions at several levels of abstraction and with regard to common problems of the social sciences and multi-agent system design, a shift in perspective that is also suggested by Castelfranchi (2000) and Panzarasa et al. (2001). In this respect Castelfranchi highlights the micro-macro link and the relationship of social functions and cognitive agents' mental representations mainly for two reasons:

First, and in particular, a theory of social function seems to be impossible to formulate without the sound knowledge of the relation between social functions and cognitive agents' mental representations, and second, social behaviour cannot be sufficiently explained without a theory of emergent social functions between cognitive (BDI) agents. Cognitive architectures are probably the most suitable way for a further analysis of this relationship. However, such an analysis also requires a sound consideration of emotion (Castelfranchi 2000: 6).

Investigation of the social functional components of emotion on three conceptual levels of analysis (micro-, meso, macro) can indeed be related to different approaches to (multi-)agent and artificial social systems quite intuitively. In these areas of inquiry, the concept of hybrid and multi-layered architectures has been brought a good step further (Castelfranchi 2000; Sloman/Logan 2000; Panzarasa/Jennings 2001/2002; Köhler/Rölke 2002). Layered architecture concepts generally build on a lower level of reactive, associative, and conditioned behaviour, on which further layers of increased deliberative capabilities and degrees of freedom rest. For example, Sloman's (2001) "CogAff" architecture is composed of reactive, deliberative, reflective, and self-conscious processes or layers, respectively.

Our approach to modelling the complex interdependencies and reciprocities of the social functions of emotion by taking a layered perspective is based on the multi-agent architecture MULAN that supplies a conceptually highly flexible framework in this respect. However, this framework is restricted by the SONAR architecture in order to achieve a clear separation of technical and conceptual models. MULAN concepts constitute a technical implementation of the agent concept that at the same time can already incorporate almost all aspects of an application model. For an improved separation of the different models we use the benefits and advantages of sociological concepts within the SONAR architecture (v. Lüde et al. 2003; Köhler et al. 2005). In particular, for each and every social unit – i.e., actors, processes, and social structures – a single SONAR agent is deployed and made available. The units' inherent logics are then described by multi-agent systems which are directly subordinated to the units in question.

In section 4 we will outline in more detail how a multi-level analysis of the social functional components of emotion can be modelled with the SONAR / MULAN architecture. But beforehand, we will further examine the role of norms and emotions in the exertion of social control.

3 Emotion and Social Control

Following the brief outline of the social functions of emotion, this section illustrates our perspective on social norms and their interactions with emotion in view of action incentives and social control.

3.1 Action Incentives

In order to lighten up the interrelation of emotion and social norms in view of the problem of structuration in artificial and natural societies, we first have to obtain clear concepts of how social norms and individual action relate to one another. If we assume that social norms constrain action and behaviour by stigmatising some options for action as more adequate than others, then we also can assume that structuration emerges in such a manner, that certain actions under certain situational conditions are not being implemented at all, and other options for action are constantly preferred by actors in such a way that robust "structuring practices" emerge (Knorr-Cetina 1981).

In this respect, Castelfranchi presupposes that in any case social norms themselves must be somehow consciously represented in order to function as action regulators. However, the *effects* of the social

norms that are intended by a norm legislating entity do not have to be explicitly represented. Thus, the functional macro-structural effects of a social norm are intended from the viewpoint of a norm legislator, but unintended from the viewpoint of a norm-complying agent. The constrained and regulated agent only adopts the *function* of a social norm: "Normative behaviour has to be intentional and conscious: it has to be based on knowledge of the norm (prescription), but this does not necessarily imply consciousness and intentionality relative to all the *functions of the norm*" (Castelfranchi 2000: 23; italics original). This view is problematic for two reasons:

First, it implicitly assumes an intentional and omnipotent norm-issuing authority that satisfies the necessary prerequisites to establish social norms in view of their overall functional societal effects. The assumption of such a global system authority seems at least doubtful and rather pointless in an MAS context. Instead, we assume that social norms emerge (unlike laws created by global system authorities, e.g. the judiciary or a dictator) according to social evolutionary principles which might be useful to the autopoiesis of a system (cf. Horne 2001; Bendor/Swistak 2001).

Second, the question whether an agent acts consciously or unconsciously in compliance with social norms is in principle insignificant, because from a sociological point of view – which is indeed also shared by Castelfranchi – it is of paramount interest what *function* is carried out by standardised (and observable) behaviour. However, the question concerning the degree of consciousness of standardised (i.e., norm-abiding) behaviour only then arises, if one investigates the mechanisms which lead to the fact that social norms despite all cognitive competence, despite intentionality, and despite a supposed free will similarly cause the same observable behaviour. An explanation for this "foundational theoretical problem of the social sciences – the possibility of unconscious, unplanned emergent forms of cooperation, organisation and intelligence among intentional, planning agents" (Castelfranchi 2000: 5) (which is also known as Adam Smith's "invisible hand") is vainly looked for solely in the area of *conscious* norm oriented behaviour. Instead, we are convinced that social norms can in fact guide human behaviour without ever becoming consciously represented, e.g. in the form of imitative or habitualised behaviour.

This does not mean, however, that norms as such are per definitionem non-propositional entities that cannot be represented consciously. In another article together with Rosaria Conte, Castelfranchi champions the idea that "*norm-abiding* behaviour need not be based on the *cognitive processing of norms* (it might be simply due to imitation)" (Conte/Castelfranchi 1995: 187; italics original). If one presupposes that cognitive processing in this respect is meant to be conscious processing, then norm-abiding behaviour can in fact occur without conscious access to the norm. This contradicts Castelfranchi's statement concerning the functions of norms, namely that "norms, to work as norms, cannot remain unconscious in the addressee: the agent should *understand them as prescriptions and use them as such*" (Castelfranchi 2000: 23; italics original).

Nevertheless, we do strongly advocate the view that, regardless of the intentional nature of social norms, they have action regulating effects by way of their attachment to emotions, especially the social emotions. However, in doing so, we emphasise that the binding of norms (or of normative behaviour) to (social) emotions is a process that largely operates on an unconscious level. Here, the norm as such remains a sub-symbolic category, an action-script whose execution is, among other factors, fostered by accompanying emotions. In fact, these sub-symbolic mental structures may under certain circumstances be explicated and communicated to others and thereby become *intentional* and *social* objects, but they by no means have to, in order to be socially functional (cf. von Scheve/Moldt (2004) for details).

To further depict the interactions between norms and emotions it seems promising to conceptualise social norms as "mental objects", and thus primarily take into account their mental, cognitive, and emotionally decisive, but not necessarily conscious components. If we follow up this approach, social norms on the one hand become instances of the macro level because of their social, temporal, and spatial distribution. On the other hand, social norms simultaneously are instances of the micro level, because they are defined as properties or configurations of propositional attitudes (i.e., beliefs, desires, intentions) and cognitive representations and have profound influences on decision-making and action selection (cf. also Engel (2002), Carley (1986/1989), and Heckathorn (1989)).

This position of norms as mental objects is analogously put forward by Conte and Castelfranchi (1995: 192) who define social norms as "hybrid configurations of beliefs and goals". According to them, social norms, being directives or instructions which are represented as *beliefs*, substantially determine future actions of an agent by generating new goals*:* "they represent a powerful mechanism for inducing new goals in people's minds in a cognitive way" (Conte/Castelfranchi 1995: 189). However, the decisive questions in this respect, „how and why does a normative belief come to interfere with *x*'s decisions? What is it that makes her [an actor] responsive to norms concerning her? What is it that makes a normative belief turn into a normative

goal?" (ibid. 192) are not answered satisfactory by Conte and Catselfranchi.

However, a sound answer to this question is of paramount significance if one wants to find a solution for the "foundational theoretical problem" mentioned by Castelfranchi. It is our conviction that emotions – especially in view of Jon Elster's (1996/1998) concept – are of outstanding importance in this respect. In the following analysis, Elster's concept shall serve as an addition to Conte and Castelfranchi's position, since Elster rather delivers a definition *of certain qualities of social norms* than of the concept of social norms itself. Correspondingly, social norms can be described as follows (Elster 1999: 145f; see also Staller/Petta 2001):

1. Social norms are non-outcome-oriented phenomena. They can have unconditional imperative character but also conditional if they refer to past actions.
2. Social norms are shared with other members of a society or a social unit in which the process of sharing itself is also socially shared.
3. The third results from the second quality, namely that behaviour in compliance with a norm is subject to enforcement by other members of a social unit, also by means of sanctions (in order to achieve the definitional social sharing).

The following section shows how far social norms influence agents' actions according to this position and which further-reaching emotion-related determinants of social action exist.

3.2 Control of Social Action

For the approach proposed here it is critical to examine the type of sanctioning in case of non-compliance to social norms. Particularly in economic theory, sanctions resulting from non-compliance are described as a withdrawal of material resources (Becker 1976; Axelrod 1986; Coleman 1990; Elster 1989). Material resources, however, are by no means the definitive or most influential objects of sanctions. Even more decisive in this respect is the fact that deviant agents interpret material sanctions also as a vehicle for the expression of negative emotions such as contempt, disdain, detestation, or disgust, and in consequence feel shame and/or guilt.

Shame in most cases will be interpreted as even worse because – in contrast to guilt – the perspective of the sanctioning agent is much more incorporated and accounted for. Furthermore, shame indicates a threat to an agent's social bonds (Scheff 2003). Elster in this context explains that the *material* aspect of sanctions lies solely within the question of how much it costs the *punisher* to impose the sanction, and not on the question of how severe the sanctions are for the offender (Elster 1999: 146). To clarify: The higher the costs a punisher accepts to implement the intended sanctions, the more insistently aware is the offender of the negative emotions lying within these sanctions, and the more strongly the offender will feel the consequent shame. The amount of the punisher's costs for sanctioning therefore signals to the offender the severity of the deviant behaviour. In many cases, punishers accept enormous costs that outreach by far the "damage" an offender has caused. But this surplus is by no means futile, since it is a way of making obvious the negative emotional meaning that comes along with the sanction and emphasises that the offender is expected to feel guilt or shame.

Frijda (1986) also takes a similar view at the interaction of social norms and emotion. He describes social rejection that results from emotional sanctions by means of shame or contempt as „severe punishment, […] most likely not merely because of its more remote adverse consequences" (Frijda 1986: 351; Elster 1999: 147). Now, what consequences do these deterrents have for agents' options to act?

Striving for emotional gratification, i.e. the motivation to seek encounters and interactions resulting in positive emotions and to avoid those resulting in negative emotions is considered a basic motivation of human behaviour. For example, Turner (1994) assumes that anxiety is one of the six primary motivational systems, whereas he defines anxiety as the "need to avoid a sense of disequilibrium with the environment" (Turner 1994: 21). Emotions – in particular anxiety – can well relate to future actions by substantially affecting their actual planning. Giddens, for example, considers concerns over a loss of ontological security to be one central aspect in his theoretical framework: it is primarily the fear of the loss of ontological security and of facticity which serves as the central motivation of action (Giddens 1991). Other authors, e.g. Collins (1984) or Hammond (1991), who think of emotional gratification as a motivator of action that is directly scalable toward social aggregational contexts, assume that actors have an inborn need for positive emotional exchange processes, which may solidify to "interaction ritual chains" and contribute to the emergence of social structures (Collins 1981; Collins 2004).

Due to the interactions of emotion and social norms explained in the previous section, we can now further assume that in particular the emotions of shame and contempt serve as vehicles for the maintenance of social norms by generating *normative goals* ("n-goals", as suggested by Conte and Castelfranchi (1995)) on the one hand and goals of avoidance of adverse consequences on the other hand. The goal of compliance with social norms therefore is *not* necessarily generated as a consequence of the anticipation of a loss of material re-

sources through sanctions, but instead as a result of the fear of emotion-driven sanctions (by means of negative emotions such as, e.g., contempt, disdain, detestation, or disgust) that again result in negative social emotions, e.g., shame, guilt, or embarrassment in the offender.

However, the significance of emotion for the structural dynamics of social systems should not solely be clarified by pinpointing to the interactions with social norms. The significance of emotion for the control of social action and also for decision-making processes is just as decisive, without normative goals being necessarily generated. This point of view is also of importance for further-reaching examinations of the role of emotion in the emergence of social norms, which, however, cannot be done in this contribution (see von Scheve/Moldt 2004 for details).

4 Modelling of Emotional Agent Systems

We start with the presentation of some aspects of emotion that have been considered significant for computer science. We basically identify three different tendencies in modelling emotions within computer systems. Subsequently, we outline the role of norms and emotion in our modelling framework SONAR, whereas the model itself is being significantly developed by specific observations of emotion as a modelling subject.

4.1 Emotional Agents

The idea to improve artificial intelligence systems by taking into account emotions or functionally equivalent mechanisms is not new; it has its origins in the contributions of authors like Simon (1967), Sloman and Croucher (1981) or Minsky (1986). In the late 1980s first reviews of existing AI-models of emotions appeared (Dyer 1987; Pfeifer 1988). Until now, research on emotion within computer science has revealed three basic motivations to equip agents with artificial emotions: performance, human computer interaction, and simulation (Wehrle 1998; Picard 1997; Scheutz 2002).

So far, research on emotional agents has largely been concerned with either isolated entities or dyadic interaction settings (agent-agent / agent-user). In view of emotional agents being applied to distributed or multi-agent systems, we suppose that the foundational functional components of emotion (artificial and/or hybrid) in social aggregates, i.e., societies, teams, groups and organisations, have to be taken into account.

First efforts in this area have already treated the role and the potential of emotions in multi-agent systems, regarding problems like structuration, coordination, cooperation, and social control (Elliot 1992; Aubé/Senteni 1996; Gmytrasiewicz/Lisetti 2000; Staller/Petta 2001). Sociological research on emotion, which primarily investigates just these very problems, could contribute to considerably extend and optimise these approaches.

Almost all computer science models and systems that include emotions share the characteristic that they are based on psychological (and neuroscientific) theories. However, since emotions bear fundamental social components and significantly influence the social phenomena which are especially interesting to distributed artificial intelligence, a sociological consideration of emotions can open new and promising perspectives for computer science and at the same time also for the social sciences (cf. Sawyer 2003; Müller et al. 1998). It seems to be most debatable to either ignore social components of emotions in AI-systems or to insufficiently consider emotional effects on social phenomena and vice versa, which get especially relevant in distributed systems.

As far as computer science claims to consider those preconditions and consequences of emotions that are constitutive for the natural phenomenon and potentially serve the purposes of computer science, it cannot possibly miss to consider also the social functions of emotion.

4.2 Emotion and Norms in SONAR

Köhler and colleagues (2003) provide a modelling framework that allows conceiving social entities, i.e. social actors, social processes, and social structures, as “first-order objects” that can be modelled “side by side” simultaneously. The framework allows complete representations of direct interdependencies which are situated on the same layer of observation / abstraction. Internal properties of a particular social entity may stay entirely encapsulated from direct access by other social entities of the same layer.

Internal logics of each entity may differ from one another significantly – they are autonomous for each social entity. This way, any actor can, e.g., have an arbitrarily complex image of its (social) environment: this image is in turn depicted through a set of networked social entities, in this case represented by means of a multi-agent system based on SONAR. The system can be very simple for primitive agents, however, it can become highly complex for agents exhibiting higher degrees of social differentiation.

For example, imagine looking at the mind of some sociologist and his internal representation of the external and internal world in its entire complexity – including all contents and any probable inconsistency. This would be a theoretical example with-

out claiming to model a real person; rather, the designer can arbitrarily simplify or extend the model according to the requirements of the task. The same is also true for the processes that are established between social actors and agents, as well as for fixed, but in the long run alterable social structures. Therefore, social entities in this framework *in themselves* contain the necessary references to other entities, i.e. these references do not have to be modelled separately.

To design a specific system, the designer is advised to deliberately choose specific aspects of a system which are supposed to be central and most relevant for the modelling goals. The SONAR models then specify exactly the most important elements of the chosen model. On the other hand, MULAN so far provides so far the technical framework for implementing the principal concepts of agents and multi-agent systems like autonomy, mobility, cooperation and adaptivity. The SONAR architecture enables modelling of the internal representation of actors in terms of a multi-agent system. Furthermore, micro-, meso- and macro-layers are being modelled, whereas micro is to be understood as an actor, meso as an interaction or process, and macro as a social structure. Figure 1 illustrates the interplay of two autonomous social entities, in this case an actor and the social process an actor is involved in. The incorporation into a social structure takes place by means of the process which is involved in the stabilisation and reproduction (or development) of new structures. Conceptually, we have to deal with the same sort of connection between both parts of the model (synchronous channels).

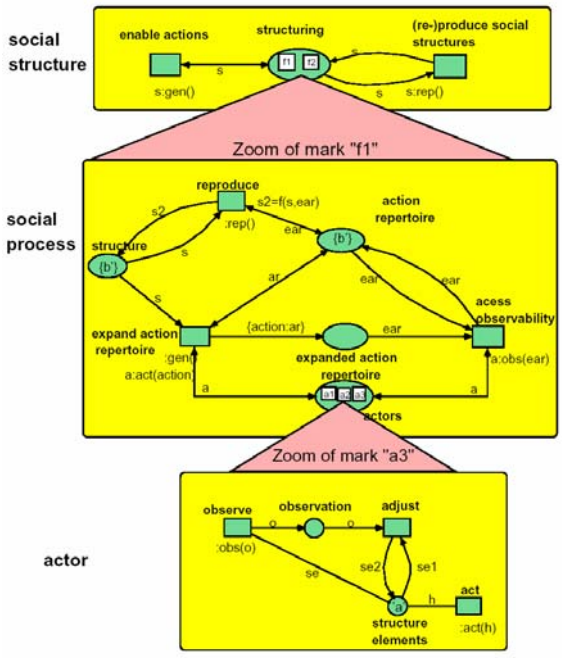

Fig. 1: A reference net representing actor, social process, and social structure (cf. v. Lüde et al. 2003)

In figure 1 a powerful variant of Petri nets – a reference net (Kummer 2002) – is shown. Rectangles (transitions) represent activities or actions, while circles (places) denote resources that can be available or not, or conditions that may be fulfilled. Arcs determine the specific context of the transition. Thus, arcs that are directed from transitions to places can be interpreted as preconditions for actions, whereas arcs that are directed from places to transitions represent actions' outcomes. A transition that fires (or: an action that is carried out) removes resources or conditions (for short: tokens) from places and inserts them into other places. A peculiar property of reference nets is the possibility that tokens on the places of a net can also be reference nets again (or either arbitrary Java-objects).

Surrounding nets are called *system nets*, those on the places *object nets* (Valk 1998). Object net and system net are synchronised by means of synchronous channels, whereby one of the nets is expected to *reference* the other one. For example, actions of an actor are synchronised with a social process by means of two transitions that constitute the synchronous channel: *observe* and *act* in actor net, here modelled as object nets, and *access visibility* and *expand action repertoire* in the social process net, in this case the system net for the actor. Tokens a1, a2 and a3 can be refined into actor nets.

It should be pointed out that this kind of refinement succeeds by means of zooming that is performed on references to the relevant actor nets. Any social entity is for itself essentially autonomous. The mutual relationship can be principally seen as symmetrical, even though modelling can imply a (rational) hierarchical structure. The concepts of respective observation and action enter the interaction via the social process entity. The adjustment succeeds locally, without direct interaction with the environment. The inscriptions of social process signify the embedding into the social structure, though these more special aspects had to be omitted here. Details can be found in either v. Lüde et al. (2003) or in Köhler et al. (2005).

Although sociological terms, such as acknowledgement, observation, action, actor, etc., are widely in use within SONAR models, the role of emotions has hardly been explicitly taken into account so far. This proves to be obstructive for modelling emotions and norms in the context of SONAR. Hence we propose some simple but fundamental enhancements to SONAR.

We can still conventionally represent the standard approaches; however we are going to treat all aspects concerning emotion separately. This approach demands an explicit decision to be made by the model designer in order to precisely determine what can and has to be classified as belonging to an emotion proper. Hence, we build a separate emotion

model besides the conventional view which is rather rational and utility-based.

The problem of connecting the separate models will be solved in a simple and homogenous manner, based on the applied modelling techniques: existing models are to be completed with further emotion models. The statements of the sociological (and other) emotion theories either deal with pure emotion-based interdependences, in which case they are to be integrated only in the emotion models, or emotions are additionally claimed to influence social structures and processes. In the latter case, linkage to the rational elements of the model is accomplished by means of synchronous channels, thus creating some kind of corresponding "parallel world". Modelling all social entities as individual nets allows arbitrarily setting them in relationship to each other through references (in the sense of a "pointer"). Through synchronous channels any possible linkage of actions from different parts of the model can be represented.

Conceptually, we therefore reach a definite separation of emotion from the so far non-emotional elements of the sociological models. Thus, emotion can be explicitly split down into its relevant integral components. Obviously, the general system complexity cannot be overcome by these means although it is distributed to different levels. Their integration demands even further efforts, since the linkages between the different levels have to be created explicitly (Here we can, e.g., proceed by applying solutions of analogous problems which can be found, for example, in the combination of different viewpoints in models created with the unified modelling language (UML).) Although at some point in the modelling process integration is necessary and essential, we currently favour the advantages resulting from an explicit separation of the models, in particular the intuitive simplicity that comes along with these detached points of view. These advantages can be clarified by referring to UML models of highly complex systems.

Smaller and less complex models can often be implemented faster with those programming languages that do not distinguish between different points of view. Though, the separation of different points of view in order to handle system complexity is a common method used in computer science. Apart from complexity issues in computer science, also sociological models are highly complex thought systems that may demand even more requirements in view of flexibility as this is generally the case for construction-oriented computer science models. Therefore, a separation into different layers seems to be essential not only for sociological emotion theories but also in view of a sharper examination of different analytical layers.

To summarise: While SONAR incorporates a division into actor, social process, and social structure, and at the same time facilitates corresponding points of view (while also operating with special model patterns, e.g. *actor* is supposed to involve observation, adoption and action (see figure 1)), we have proposed a further dimension which is especially significant because of its content-based and therefore application-related qualities: the supplementary modelling of emotion.

The possibility to design discrete emotions and their specific components as explicit states or processes which are integrated into the existing (rational) models still remains unelaborated, but is a goal for future work. Although the approach presented here allows a sharp distinction, it does not enforce it, since profound experiences with applications are still missing and restrictions of the flexibility might be possible. However, these restrictions should only arise from the requirements of the specific theories of emotion, which can provide the appropriate and substantial arguments.

5 Discussion

Finally we briefly evaluate the modelling approach presented above as well as the possibilities it bears in view of modelling the interrelation of emotion and norms:

Generally, by considering emotions the capabilities of simulating and modelling sociological theory are crucially improved. A differentiated presentation facilitates the reduction of complexity concerning the adopted view and the particular system, supported by the capacities of UML. The technical implementation of emotional mechanisms within the SONAR / MULAN-architecture is still under current development.

This simulation environment can be subsequently used to improve multi-agent systems, for example in view of alternative coordination solutions. However, evidence must yet be presented, whether modelling emotions can ever enable such solutions. Still, following Minsky (1986), we take it for granted, that intelligent systems need a replication of the (social) functions of emotion.

The interaction between emotion and deliberation/cognition can now be investigated on the basis of explicitly separated components of a model, whereby specific sociological questions can also be addressed. The acknowledged separation primarily aims at analytical clarity and the explicit modelling of interactions and interdependencies, and it does not target the explanation of natural phenomena.

The explicit representation of emotion on the basis of analytical models that are still to be developed

can in addition be used and further scrutinised in the context of human-computer interaction.

The usage of SONAR with its conceptual distinction of actors, processes, and structures as well as patterns, that are to be specified still more precisely, makes it possible to reflect the central social theoretical concepts, such as the interrelation of norms and emotion, especially in the contexts mentioned above.

There exists a further architecture that is directly embedded into a FIPA-conform agent-system framework (CAPA) (Duvigneau et al. 2003). Its conceptual foundation MULAN provides software-technical support of application-related concepts, thus sparing the transfer efforts of the model designer.

The dependences between norms and emotions can now be examined on different levels. Thus, on the structural layer we can combine norms either with the modelling concepts, which largely disregard emotions, or with those that consider emotions on all layers of a model (actor, process, and structure).

This flexibility permits to consider different theories of emotion *simultaneously* and to different extents. At this point, the existing emotional agent architectures, which are primarily concerned with modelling actor- and process-layer (e.g., TABASCO (Staller/Petta 1998)), can be picked up, integrated, and extended with the sociological aspects presented above.

The decomposition into actor, process, and structure also supports the constitution of the examined emotion theories that expand from the neurological and cognitive to the sociological points of view. Moreover, this separation may crucially simplify the formulation of an interdisciplinary theory of emotion, since it acknowledges the interconnections between sociological, psychological, and probably also neurological theories (see von Scheve/Moldt 2004).

Acknowledgements

The first author acknowledges a grant from the Research Group “Emotions as Bio-Cultural Processes“, Center for Interdisciplinary Research (ZiF), University of Bielefeld, Germany.

References

J.A. Abe and C.E: Izard. The Developmental Functions of Emotions: An Analysis in Terms of Differential Emotions Theory. *Cognition & Emotion*, 13(5): 523-549, 1999.

J.R. Averill. The Structural Bases of Emotional Behavior. In: M.S. Clark (Ed.). *Emotion*. Review of Personality and Social Psychology, Vol. 13. Newbury Park/CA: Sage, 1-24, 1992.

M. Aubé and A. Senteni. Emotions as Commitments Operators: A Foundation for Control Structure in Multi-Agents Systems. In: W. van de Velde and J.W. Perram (Eds.). *Agents Breaking Away*. Proceedings of the 7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW’96), January 22.-25., Eindhoven/NL. LNAI Vol. 1038. Berlin: Springer, 13-25, 1996.

R. Axelrod. An evolutionary approach to norms. *American Political Science Review*, 80: 1095-1111, 1986

A.L. Bazzan, D.F. Adamatti and R. Bordini. Extending the Computational Study of Social Norms with a Systematic Model of Emotions. In: G. Bittencourt and G.L. Ramalho (Eds.). *Advances in Artificial Intelligence*. Proceedings of the 16th Brazilian Symposium on Artificial Intelligence (SBIA’02), Porto de Galinhas/Recife, Brazil, November 11-14. LNCS Vol. 2507. Berlin: Springer, 108-117, 2002.

J. Bendor and P. Swistak. The Evolution of Norms. *American Journal of Sociology*, 106(6): 1493-1545, 2001

G. Becker. *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press, 1976.

J.T. Cacioppo, G.G. Berntson, J.T. Larsen, K.M. Poehlmann and T.A. Ito. The Psychophysiology of Emotion. In: R. Lewis and J.M. Haviland-Jones (Eds.). *Handbook of Emotions*, 2nd Ed. New York: Guilford Press, 173-191, 2000.

D. Cañamero (Ed.). *Emotional and Intelligent: The Tangled Knot of Cognition*. Proceedings of the 1998 AAAI Fall Symposium. Technical Report FS-98-03. Menlo Park/CA: The AAAI Press, 1998.

K. Carley. An Approach for Relating Social Structure to Cognitive Structure. *Journal of Mathematical Sociology*, 12(2): 137-189, 1986.

K. Carley. The Value of Cognitive Foundations for Dynamic Social Theory. *Journal of Mathematical Sociology*, 14(2/3): 171-208, 1989.

C. Castelfranchi. The Theory of Social Functions: Challenges for Computational Social Science and Multi-Agent Learning. *Journal of Cognitive Systems Research*, 2(1): 5-38, 2000.

G.L. Clore, N. Schwarz and M. Conway. Affective Causes and Consequences of Social Information Processing. In: R.S. Wyer and T.K. Srull (Eds.). *Handbook of Social Cognition*. 2nd Ed. Hillsdale/NJ: Lawrence Erlbaum, 323-417, 1994.

G.L. Clore. Why Emotions Are Felt. In: P. Ekman and R.J. Davidson (Eds.). *The Nature of Emotion*. New York: Oxford University Press, 103-111, 1994.

J.S. Coleman. *Foundations of Social Theory*. Cambridge/MA: Harvard University Press, 1990.

R. Collins. On the Microfoundations of Macrosociology. *American Journal of Sociology*, 86(5): 984-1014, 1981.

R. Collins. The Role of Emotion in Social Structure. In: K.R. Scherer and P. Ekman (Eds.). *Approaches to Emotion*. Hillsdale/NJ: Lawrence Erlbaum, 385-396, 1984.

R. Collins. *Interaction Ritual Chains*. Princeton/NJ: Princeton University Press, 2004.

R. Conte and C. Castelfranchi. Norms as Mental Objects – From Normative Beliefs to Normative Goals. In: C. Castelfranchi and J.-P. Müller (Eds.). *From Reaction to Cognition*. Selected Papers from the 5th European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW '93), 25.-27. August, Neuchatel/CH. LNAI Vol. 957. Berlin: Springer, 186-196, 1995.

L. Cosmides and J. Tooby. Evolutionary Psychology and the Emotions. In: R. Lewis and J.M. Haviland-Jones (Eds.). *Handbook of Emotions*, 2nd Ed. New York: Guilford Press, 91-115, 2000.

F. Dignum, D. Morley, E.A. Sonenberg and L. Cavedon. Towards socially sophisticated BDI agents. In: *Proceedings of the 4th International Conference on Multiagent Systems (ICMAS'00)*. Boston/MA: IEEE, 2000.

U. Dimberg. Psychophysiological Reactions to Facial Expressions. In: U. Segerstrale and P. Molnar (Eds.). *Nonverbal Communication: Where Nature Meets Culture*. Mahwah/NJ: Lawrence Erlbaum, 47-60, 1997.

E. Durkheim. *Die elementaren Formen des religiösen Lebens*. Frankfurt am Main: Suhrkamp, (1994)[1912].

M. Duvigneau, D. Moldt and H. Rölke. Concurrent Architecture for a Multi-agent Platform. In: F. Giunchiglia, J. Odell and G. Weiß (Eds.). *Agent-Oriented Software Engineering III*. Proceedings of AOSE'02. LNCS Vol. 2585. Berlin: Springer, 59-72, 2003.

M.G. Dyer. Emotions and their computations: Three computer models. *Cognition & Emotion*, 1(3): 323-347, 1987.

C.D. Elliot. Using the Affective Reasoner to Support Social Simulations. *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, Vol. 1. Los Altos/CA: Morgan Kaufmann, 194-200, 1993.

C.D. Elliot. *The Affective Reasoner. A Process Model of Emotions in a Multi-Agent System*. PhD Thesis, Technical Report 32, Intitute for the Learning Sciences, Northwestern University, 1992.

J. Elster. Emotions and Economic Theory. *Journal of Economic Literature*, 36(1): 47-74, 1989.

J. Elster. Rationality and the Emotions. *The Economic Journal*, 106(438): 1386–1397, 1996.

J. Elster. *Alchemies of the Mind. Rationality and the Emotions*. Cambridge: Cambridge University Press, 1999.

P. Engel. Intentionality, Normativity and Community. *Facta Philosophica* 4(1): 25-49, 2002.

J. Fix. *Emotionale Agenten*. Diploma Thesis. University of Hamburg, Computer Science Department, 2004.

N.H. Frijda. *The Emotions*. Cambridge: Cambridge University Press, 1986.

J. Gerhards. *Soziologie der Emotionen: Fragestellungen, Systematik u. Perspektiven*. Weinheim: Juventa, 1988.

A. Giddens. *Modernity and Self-Identity*. Cambridge: Polity Press, 1991.

P.J. Gmytrasiewicz and C.L. Lisetti. Using Decision Theory to Formalize Emotions for Multi-Agent Systems. *Proceedings of the 2nd ICMAS'00 Workshop on Game Theoretic and Decision Theoretic Agents*. Boston/MA, 2000.

G. Hatano, N. Okada and H. Tanabe (Eds.). Affective Minds. Proceedings of the 13th Toyota Conference. Amsterdam: Elsevier, 2000.

D.D. Heckathorn. Cognitive Science, Sociology, and the Theoretic Analysis of Complex Systems. *Journal of Mathematical Sociology*, 14(2/3): 97-110, 1989.

A.R. Hochschild. Emotion Work, Feeling Rules, and Social Structure. *American Journal of Sociology*, 85(3): 551-575, 1979.

A.R. Hochschild. *The Managed Heart*. Berkeley/CA: University of California Press, 1983.

M. Holodynski and W. Friedlmeier. *Development of Emotions and Their Regulation*. Boston/MA: Kluwer Academic, 2005[in press].

C. Horne. Sociological Perspectives on the Emergence of Norms. In: M. Hechter and K.-D. Opp (Eds.). *Social Norms*. New York: Russell Sage Foundation, 3-34, 2001.

Keltner, D.; Gross, J.J. (1999): Functional Accounts of Emotion. *Cognition & Emotion*, 13(5): 467-480.

D. Keltner and J. Haidt. Social Functions of Emotion at Four Levels of Analysis. *Cognition & Emotion*, 13(5): 505-521, 1999.

T.D. Kemper. *A Social Interactional Theory of Emotions*. New York: Wiley & Sons, 1978.

K.D. Knorr-Cetina. Introduction: The micro-sociological challenge of macro-sociology. In: K.D. Knorr-Cetina and A.V. Cicourel (Eds.). *Advances in social theory and methodology*. Boston/MA: Routledge & Kegan Paul, 1-47, 1981.

M. Köhler, D. Moldt and H. Rölke. Modeling the Structure and Behaviour of Petri Net Agents. In: J.M. Colom and M. Koutny (Eds.). *Proceedings of the 22nd International Conference on Application and Theory of Petri Nets (ICATPN'01)*, June 25.-29., Newcastle/UK. LNCS Vol. 2075. Berlin: Springer, 224-241, 2001.

M. Köhler, D. Moldt and H. Rölke. A Discussion of Social Norms with Respect to the Micro-Macro Link. In: *Proceedings of the 2nd International Workshop on Regulated Agent-Based Social Systems: Theories and Applications* (RASTA'03), June 24.-26., Edinburgh, 2003.

M. Köhler, D. Moldt, H. Rölke and R. Valk. Linking micro and macro descriptions of scalable social systems using reference nets. In: K. Fischer and M. Florian (Eds.). *Socionics: Its Contribution to the Scalability of Complex Social Systems.* LNCS. Heidelberg: Springer, 2005[in press].

M. Köhler and H. Rölke. Modelling the Micro-Macro Link: Towards a Sociologically Grounded Design of Multi Agent Systems. In: C. Jonker, G. Lindemann and P. Panzarasa (Eds.). *Proceedings of the Workshop Modelling Artificial Societies and Hybrid Organizations (MASHO'02)*, 2002.

O. Kummer. *Referenznetze*. Berlin: Logos-Verlag, 2002.

E.J. Lawler, S.R. Thye and J. Yoon. Emotion and Group Cohesion in Productive Exchange. *American Journal of Sociology*, 106(3): 616-657, 2000.

R. von Lüde, D. Moldt and R. Valk (Eds.). *Sozionik. Modellierung soziologischer Theorie*. Münster: Lit-Verlag, 2003.

D. Moldt and C. von Scheve. Emotions in Hybrid Social Aggregates. In: M. Herczeg, W. Prinz and H. Oberquelle (Eds.). *Mensch & Computer 2002. Vom interaktiven Werkzeug zu kooperativen Arbeits- und Lernwelten*. Stuttgart: Teubner, 343-352, 2002.

M. Minsky. *The Society of Mind*. New York: Simon & Schuster, 1986.

H.J. Müller, T. Malsch and I. Schulz-Schaeffer. Socionics: Introduction and Potential. *Journal of Artificial Societies and Social Simulation*, 1(3), 1998.
http://www.soc.surrey.ac.uk/JASSS/1/3/5.html

K. Oatley and J.M. Jenkins. *Understanding Emotions*. Oxford: Basil Blackwell, 1996.

P. Panzarasa and N.R. Jennings. The Organisation of Sociality: A Manifesto for a New Science of Multi-Agent Systems. *Proceedings of the 10th European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'01)*, Annecy, France, 2001.

P. Panzarasa and N.R. Jennings. Social Influence, Negotiation and Cognition. *Simulation Modelling Practice and Theory*, 10(5-7): 417-453, 2002.

P. Panzarasa, T.J. Norman and N.R. Jennings. Social mental shaping: Modelling the impact of sociality on autonomous agents' mental states. *Computational Intelligence*, 17(4): 738-782, 2001.

R. Pfeifer. Artificial Intelligence Models of Emotion. In: V. Hamilton, G.H. Bower and N.H. Frijda (Eds.). *Cognitive Perspectives on Emotion and Motivation*. Dodrecht: Kluwer Academic, 287-320, 1998.

R.W. Picard. *Affective Computing*. Cambridge/MA: The MIT Press, 1997.

N.J. Saam and A. Harrer. Simulating Norms, Social Inequality, and Functional Change in Artificial Societies. *Journal of Artificial Societies and Social Simulation*, 2(1), 1999. http://www.soc.surrey.ac.uk/JASSS/2/1/2.html

R.K. Sawyer. Artificial Societies. Multiagent Systems and the Micro-Macro Link in Sociological Theory. *Sociological Methods & Research*, 31(3): 325-363, 2003.

R.K. Sawyer. Social Explanation and Computational Simulation. *Philosophical Explorations*, 7(3): 219-231, 2004.

T.J. Scheff. Shame in Self and Society. *Symbolic Interaction*, 26(2): 239-262, 2003.

M. Scheutz. Agents with or without Emotions? In: S. Haller and G. Simmons (Eds.). *Proceedings of the 15th International Florida Artificial Intelligence Symposium* (FLAIRS'02). Menlo Park/CA: The AAAI Press, 89-94, 2002.

C. von Scheve and D. Moldt. Emotion: Theoretical Investigations and Implications for Artificial Social Aggregates. In: G. Lindeman, D. Moldt and P. Paolucci (Eds.): *Regulated Agent-Based Social Systems*. LNAI Vol. 2934. Berlin: Springer, 189-209, 2004.

N. Schwarz. Feelings as information: Informational and motivational functions of affective states. In: R.M. Sorrentino and E.T. Higgins (Eds.). *Handbook of Motivation and Cognition*. Foundations of Social Behavior, Vol. 2. New York: Guilford Press, 527-561, 1990.

H.A. Simon. Motivational and emotional controls of cognition. *Psychological Review*, 74: 29-39, 1967.

A. Sloman. Varieties of Affect and the CogAff Architecture Schema. In: *Agents and Cognition. Proceedings of the AISB'01 Symposium on Emotion, Cognition, and Affective Computing*. York: SSAISB, 39-48, 2001.

A. Sloman and M. Croucher. Why robots will have emotions. *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI'81)*. Vancouver, British Columbia, 197-202, 1981.

A. Sloman and B. Logan. Evolvable Architectures for Human-Like Minds. In: G. Hatano, N. Okada and H. Tanabe (Eds.). *Affective Minds*. Proceedings of the 13th Toyota Conference, Nov.-Dec. 1999, Nagoya/Japan. Amsterdam: Elsevier, 169-182, 2000.

C.A. Smith and L.K. Pope. Appraisal and Emotion. The Interactional Contributions of Dispositional and Situational Factors. In: M.S. Clark (Ed.). *Emotion and Social Behavior*. Review of Personality and Social Psychology, Vol. 14. Newbury Park/CA: Sage, 32-62, 1992.

A. Staller and P. Petta. Towards a tractable appraisal-based architecture for situated cognizers. In: D. Cañamero, C. Numaoka and P. Petta (Eds.). *Grounding Emotions in Adaptive Systems*. Workshop Notes of the 5th International Conference of the Society for Adaptive Behaviour (SAB'98). Zurich, Switzerland, 56-61, 1998.

A. Staller and P. Petta. Introducing Emotions into the Computational Study of Social Norms: A First Evaluation. *Journal of Artificial Societies and Social Simulation*, 4(1), 2001. http://soc.surrey.ac.uk/JASSS/4/1/2.html

R. Trappl, P. Petta and S. Payr (Eds.). *Emotions in Humans and Artefacts*. Cambridge/MA: The MIT Press, 2003.

J.H. Turner. A General Theory of Motivation and Emotion in Human Interaction. *Österreichische Zeitschrift für Soziologie*, 19(1): 20-35, 1994.

J.H. Turner. *On the Origins of Human Emotions. A Sociological Inquiry into the Evolution of Human Affect*. Stanford: Stanford University Press, 2000.

J.H. Turner. *Face to Face. Toward a Sociological Theory of Interpersonal Behavior*. Stanford/CA: Stanford University Press, 2002.

R. Valk. Petri Nets as Token Objects. An Introduction to Elementary Object Nets. In: J. Desel and M. Silva (Eds.). *Proceedings of Application and Theory of Petri Nets*. Lisbon, Portugal. LNCS Vol. 1420. Berlin: Springer, 1-25, 1998.

T. Wehrle. Motivations Behind Modeling Emotional Agents: Whose emotions does your robot have? In: D. Cañamero, C. Numaoka and P. Petta (Eds.). *Grounding Emotions in Adaptive Systems*. Workshop Notes of the 5th International Conference of the Society for Adaptive Behaviour (SAB'98). Zurich, Switzerland, 1998.