Comput Math Organiz Theor (2006) 12:81–100
DOI 10.1007/s10588-006-9538-6

My agents love to conform: Norms and emotion in the micro-macro link

Christian von Scheve · Daniel Moldt · Julia Fix · Rolf von Luede

Abstract This contribution investigates the function of emotion in relation to norms, both in natural and artificial societies. We illustrate that unintentional behavior can be normative and socially functional at the same time, thereby highlighting the role of emotion. Conceiving of norms as mental objects we then examine the role of emotion in maintaining and enforcing such propositional attitudes. The findings are subsequently related to social structural dynamics and questions concerning micro-macro linkage, in natural societies as well as in artificial systems. Finally, we outline the possibilities of an application to the socionic multi-agent architecture SONAR.

Keywords Emotions · Social norms · Micro-macro link · Multi-agent systems · Socionics · Petri Net modeling

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 consisting of a multitude of intelligent autonomous agents. Agent systems are feared to run out of control in such a way that autonomously generated goals a system pursues might contradict some of the crucial, although implicit high-level goals of the designer or the user of a system, in consequence leading to inefficient, undesirable or even hazardous system

C. von Scheve (✉) · R. von. Luede
University of Hamburg, School of Business, Economics and Social Sciences, Institute of Sociology, Allende-Platz 1, D-20146 Hamburg, Germany
e-mail: xscheve@informatik.uni-hamburg.de

D. Moldt · J. Fix
Faculty of Informatics, Theoretical Foundations of Computer Science, Vogt-Koelln-Str. 30, D-22527 Hamburg, Germany

behavior. To make matters worse, the dilemma arising out of these conflicting goals affects some of the core strengths of artificial agent systems: their autonomy, flexibility, and discretion. Thus, approaches at resolving the issue of conflicting goals have to ensure the autonomy of the system in question and the integrity of human high-level goals at the same time. One feasible solution to this problem is the implementation of a system of shared norms which is not coerced by the designer, but instead emerges from mutual interactions of agents (and users) (cf. Shoham and Tennenholtz, 1997). To support such an approach, it might be beneficial—if not mandatory—to acquire profound knowledge of and to adapt to the computational context the mechanisms of norm emergence, maintenance, and compliance in human social systems. In this article, we argue that emotion constitutes such a mechanism and should therefore be taken into account in designing agents and multi-agent systems (MAS) in particular.

The concept of *agents* is inspired by and to a large extent relies on classical paradigms of cognitive psychology and hence on corresponding conceptualizations of intelligent behavior, which is still fundamentally based on cognitions and representations. Belief-Desire-Intention (BDI) architectures can well be considered an epitome of this perspective on intelligent information processing (Wooldridge, 2000). Notwithstanding this, the interrelation of emotion and cognition and the role of emotion in overall intelligent behavior have been long debated in the behavioral sciences, having likewise promoted the idea that artificial intelligence (AI) systems might be improved by taking into account mechanisms which are functionally equivalent to emotions in their biological counterparts (Simon, 1967; Sloman and Croucher, 1981). At least since Marvin Minsky's programmatic and frequently evoked 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. entities endowed with mechanisms functionally equivalent to emotion in human and non-human animals (cf. Canamero, 1998; Hatano, Okada, and Tanabe, 2000; Trappl, Petta, and Payr, 2003).

Yet, the functions of emotion in large-scale distributed systems have not been investigated extensively, despite some initial explorative efforts in multi-agent systems research (Aube and Senteni, 1996; Elliot, 1993; Fix, 2004; Gmytrasiewicz and Lisetti, 2000; von Scheve and Moldt, 2004). This is probably due to fact that research on emotions in natural large-scale social systems, i.e. human societies, groups, organizations, and communities is still largely underdeveloped, although first investigations indicate that emotion indeed plays a key role in interlinking individual action and global system behavior (Barbalet, 1998; von Scheve and von Luede, 2005). A crucial part of these initial efforts is the analysis of the interplay of social norms and emotion. Proper theorizing and empirical evidence both suggest that emotion is essential in sustaining social norms and in enforcing sanctions in cases of non-compliance (Elster, 1996; Fehr and Fischbacher, 2004). On the other hand, social norms are frequent causes of emotions and to a certain extent determine coping and regulation processes (Heise and Calhan, 1995; Hochschild, 1979).

Whereas the application of norms as a determinant of social action and a principle of coordination is becoming a major area of inquiry in multi-agent systems research and may draw on an extensive literature in the social sciences (Castelfranchi et al., 1999; Verhagen, 2000; Tuomela, 1995), it is not surprising that the role of emotion in normative multi-agent systems is largely not accounted for. However, pioneering research in the computational study of social norms and emotion has been conducted by Alexander Staller and Paolo Petta (2001) and by Ana Bazzan and colleagues (2002). In this article, we aim at extending this line of research by following a socionic approach, considering agent systems as a model for social science research—in fact a modeling tool—and a software engineering paradigm alike (Jennings, 2000; von Luede, Moldt, and Valk, 2003; Macy and Willer, 2002; Malsch, 2001; Sawyer, 2004). In view of the fact that distributed systems are of increasing importance in many areas of application, for example in electronic marketplaces, automated negotiations, planning and scheduling systems, business process and workflow management, coordination of large-scale open systems, and social simulations it seems thoroughly reasonable to us to further promote the investigation of the functions of emotion in large-scale social systems, natural as well as artificial.

We proceed as follows: First, we briefly illustrate some of the social functions of emotion in a broader perspective regarding an organism's internal functioning and social interactions. We then argue for a model of social control that is fundamentally based on the relation between social norms and emotion, thereby referring to two specific functions of emotion and the nature of social norms. In this model, the emotion-based commitment to social norms in particular ensures actors' compliance with norms. Finally, we give a sketch of how these findings might be modeled in the SONAR/MULAN multi-agent architectures.

Social functions of emotion

To achieve a better understanding of the relation between social norms and emotion, we will outline some of the social functions emotion serves in individual behavior, social interaction, and social aggregation, which may in turn be distinguished from intraindividual (Levenson, 1999), phylogenetic (Cosmides and Tooby, 2000), and ontogenetic functions (Abe and Izard, 1999) of emotion. In fact, emotions are supposed to be functional *by definition*, being regarded as "functional, organized responses to environmental demands that prepare and motivate the person to cope with the adaptational implications of those demands" (Smith and Pope, 1992: 36). According to this definition, a central function of emotion is the adaptational and beneficial regulation of behavior in relation to environmental conditions, which clearly can be both, physical and social in nature (Keltner and Gross, 1999: 468). Conceptually as well as empirically, social functions of emotion have been located on different levels, for example biological, psychological, and social (Averill, 1992); individual, dyadic, group, and cultural (Keltner and Haidt, 1999); organism, personality, social structure, and culture (Gerhards, 1988) as well as micro, meso, and macro (von Scheve and Moldt, 2004).

On the level of an individual agent, emotion performs above all two functions: On the one hand, emotion informs an agent about events in the social environment that require immediate, reactive, and adaptive behavior, also known as the "affect as information" paradigm (Clore, Schwarz, and Conway, 1994; Schwarz and Clore, 1988). 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 (Keltner and Haidt, 1999; see below). On the other hand, emotions prepare the organism to react adequately upon situational demands, for example by triggering physiological arousal and by structuring the cognitive system into adequate operational modes (Cacioppo et al., 2000; Clore, 1994; Oatley and Jenkins, 1996: 252). The verbal and nonverbal expression of emotion in social interaction is a further function of emotion particularly tied to social norms (Keltner and Haidt, 1999): First of all, emotion expressions allow the mutual attribution of most interactional contingencies, including emotional state, interpretations of the situation, and intentions. Secondly, emotion expressions may (unconsciously) evoke complementary or reciprocal reactions in context-bound observing actors and therewith contribute to improved mutual interpretations of a situation, which in turn is a prerequisite for cooperation and the coordination of social action (Frank, 2001). Thirdly, emotion expressions may promote or constrain specific courses of an interaction and act as either motivating or sanctioning signals. Large-scale perspectives on the social functions of emotion highlight their role in identifying social groups and group members (Heise and O'Brien, 1993), ascribing status and power resources (Kemper, 1978), the emergence and maintenance of solidarity and cohesion (Lawler, Thye, and Yoon 2000), and in the internalization and sustenance of social norms, power structures, moral and cultural ideas, ideologies, and the like (Barbalet, 1998; Elster, 1999; Hochschild, 1979).

As outlined above, social functions of emotion are identified on rather different levels, all of which supposedly "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", to stick to Keltner and Haidt's (1999: 507, italics added) view on functional explanations of emotion. The use of different levels of abstraction in analyzing the social functions of some trait or behavior is also common in general social theory and in multi-agent systems design (cf. below) and has proven to be particularly helpful in micro-macro analyses of social structural dynamics (Lawler, Ridgeway, and Markowsky, 1993; Sawyer, 2003; Wiley, 1988). In discussing Elster's notion of functional explanation, Castelfranchi (2001) convincingly argues that in order to elucidate what is functional for a social system the relationship of functional behavior and cognitive agents' mental representations has to be thoroughly understood. Cognitive (BDI) architectures are probably a suitable tool for the further analysis of this relationship, even though such an analysis requires the profound consideration of emotion, as Castelfranchi (2001: 11) casually indicates and we will elaborate in more detail now.

In an artificial intelligence context, the major problem arising out of this perspective on the social functions of emotion seems to be a physiological and biological rather than a cognitive one and is indeed a problem almost every disembodied agent system has to face. Although the cognitive aspects of emotion (and vice versa) are far from being thoroughly understood—let alone computationally modeled—some of the above mentioned theories, as well as our own interdisciplinary approach (von Scheve and von Luede, 2005), highlight the role of physiological processes in analyzing the functions of emotion. These approaches stand in sharp contrast to other social constructionist perspectives on emotion that largely deny any importance of physiology. Simon Clarke (2003) has aptly outlined this dilemma in the case of envy, thereby emphasizing the importance of psychoanalytic theory and the limitations of a purely social constructionist, i.e. cognitivist, account particularly in explaining the role of emotion in large-scale behavioral dynamics.

Given these issues, there are, nevertheless, some reasons we are optimistic that such an endeavor might indeed be successful: First is the fact that models of emotion that encompass the physiological level are just beginning to emerge (e.g., Avila-Garcia and Canamero, 2005; Mendao, 2004; de Almeida, da Silva, and Bazzan, 2004) and might well contribute to more sophisticated multi-level theories also incorporating the physiological level. Second, the role of cognition in emotion has been exhaustively accounted for, but we still see the potential of further refinement of existing models. Third, we are foremost concerned with modeling interactions and communication *between* the different layers, and might therefore duly leave some of the aspects temporarily black-boxed.

Normative behavior and emotion

In many, if not most, normative multi-agent systems norms are conceptualized with reference to deontic logic as explicitly represented obligations, duties, and conventions determining "that something ought or ought not to be the case" under specific circumstances (Opp, 2002: 132). In addition, norms are characterized by various features, e.g. their acceptance, maintenance, social distribution, formalization, rationale, function or purpose (Verhagen, 2000). Whatever aspect is highlighted in the various approaches to normative multi-agent systems, many entities in these systems are assumed to be "norm autonomous," i.e. may autonomously and intentionally decide to follow or violate a norm, supposed that they have full knowledge of actually prevailing norms in a social system and a specific social situation and perfect information on their internal milieu and external environment. Furthermore it is often assumed that agents are able to deliberately and collectively issue norms (Castelfranchi et al., 1999; Dignum, 1999).

This perspective on deliberative normative agents raises a couple of issues, in particular when dealing with problems of micro-macro linkage. We will show that actors are not as "autonomous" as they might seem when it comes to norms, and that other forms of normative behavior are equally important when dealing with micro-macro issues. First and foremost is the problem of deliberate choice: If an agent may always autonomously decide either to follow or to violate a norm, then wherein lies the purportedly outstanding and compulsive power of norms and on what basis are such decisions derived? In highly functional and institutionalized domains such as markets, law, or politics, decision-makers usually act rationally in order to maximize individual or collective outcomes. Given a rational normative agent, its decision to adhere to a norm will therefore be guided by the subjectively expected outcome of the different alternatives. However, there are good examples where abiding to a norm is highly irrational in terms of costs and benefits, e.g. in vendettas, altruistic punishment, tax compliance, or voting in large-scale elections. Nevertheless, most actors still do follow these norms (Elster, 2004). Another example is the fact that people follow norms even in solitude and in absence of probable sanctions (including loss of reputation), for example in adhering to table manners or by contributing to the provision of public goods (e.g. protection of the environment).

In addition, the problem of deliberate choice is particularly evident when it comes to what is commonly understood as *social* norms, i.e. norms that are not only socially shared but in addition involve third-party punishment as a defining feature and a normative action in itself. Evidence suggests that third-party punishers, which are by definition materially unaffected by norm violations, carry out sanctions even at their own costs with no immediate benefits (material, reputation, deterrence) (Fehr and Fischbacher, 2004a). Why do they choose to do so? It has also been shown that third-party punishment, i.e. the existence of a social norm prescribing the punishment of violators, promotes evolutionary more stable exchange strategies in iterated dyadic interactions than, for example, second-party punishment (Bendor and Swistak, 2001).

This brings us to another issue, namely the emergence of norms. If norms are to be intentionally issued for instrumental, i.e. socially functional reasons, for example to promote the fitness of some large-scale social system, then it seems highly improbable that there exists an individual or corporate norm-legislating entity capable of determining what is in fact most functional for such a large-scale system, thereby simultaneously accounting for different units of measuring utility, differing time-scales, imperfect information, etc. The intentional issuing of norms seems to be practicable in small to medium-scale social systems, e.g. organizations and communities, but highly impracticable as the size and complexity of a system increases (cf. Bendor and Swistak, 2001: 1495). Many norms are not issued intentionally but rather emerge from the unintended consequences of actions, for example table manners, dress codes, and fairness rules (cf. Opp, 2002). Given these issues, some researches in the social and behavioral sciences still embrace the idea of explaining the social functions of norms solely by referring to their rational qualities (Opp, 2002; Horne and Cutlip, 2002; Bendor and Swistak, 2001), while others—such as us—do not (Elster, 1996, 2004; Frank, 1993, 2001).

Robert Axelrod in a seminal article on the evolution of norms uses a behavioral definition of norms: "A norm exists in a given setting to the extent that individuals usually act in a certain way and are often punished when seen acting not in this way" (Axelrod, 1986: 1097). Although Axelrod applies this definition to a game-theoretic context, we will use it to suggest a behavior-based approach to normative agents and to explain the functions of norms and emotions in a micro-macro context. This does not mean, however, that we categorically disregard the possibility of explicitly represented norms. All we aim at is the distinction between an explicit representation of a norm on the one hand, and the existence of unintentional normative behavior that is not based on some propositional mental content. A norm may well emerge from initially uncoordinated and then for some reason (or social function, if one wants) successively increasing cooperative behavior, and later on (if at all) become explicitly represented, for example as a formal law. Or, on the other hand, a norm is intentionally formulated, issued and adopted, and thus promotes normative behavior.

If we assume that social norms constrain actions by denoting some options for action as more adequate than others, then it is also plausible to assume that social structuration emerges in such a way that certain actions under certain situational conditions are not being implemented at all, while other options are constantly preferred, leading to the emergence of robust "structuring practices" (Knorr-Cetina, 1981). In this respect, Cristiano Castelfranchi (2001) gives a good account of the interrelation of norms and social functions (and, casually, emotion). Although his definition of norms is not in line with our concept, he develops an interesting stimulus-reinforcement learning model to account for structural effects of individual behavior. In addressing the "foundational theoretical problem of the social sciences—the possibility of unconscious, unplanned emergent forms of cooperation, organization and intelligence among intentional, planning agents" (Castelfranchi, 2001: 6), Castelfranchi suggests that the reinforcement of goals and beliefs as the main determinants of social action in effect leads to the reinforcement of corresponding goal-directed behavior. In doing so, he postulates two central mechanisms, namely the strengthening of associations between beliefs/goals and situational contexts on the one hand, and the confirmation of beliefs supporting a desired behavioral outcome on the other hand.

Although Castelfranchi criticizes Bourdieu's "habitus" for being far too deterministic and lacking an explanation of the social functions of intentional action, his model of reinforcement learning at least to us strongly resembles the very characteristics of habitual behavior (Castelfranchi, 2001: 25f). As far as we can assess, Bourdieu did not favor an opposition between intentional behavior as non-functional and the habitus as being socially functional. On the contrary, empirically—and this relates to a general problem with Castelfranchi's account and a common misunderstanding of Bourdieu's habitus concept (see Lizardo (2004) for an excellent clarification on this matter)—*there is no such thing as "pure" conscious intentional behavior devoid of any social influence or habitualization* (cf. von Scheve and von Luede, 2005). Notably, Castelfranchi in his layered architectural approach argues for exactly this interaction of associative, habitualized low-level mechanisms and high-level cognitive functions, in that he proposes an additional layer of automatic, associative mechanisms resting *upon* the layer of explicit mental representations, propositional attitudes, and higher cognitive functioning. Although this kind of behavior is considered social functional (i.e. structurally reinforcing), Castelfranchi vividly denies its *normative* status. "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, 2001: 31; italics omitted). In contrast to this view, we argue that the potential of social norms in explaining social structural dynamics is vainly looked for solely in the area of conscious and intentional action. If one conceives of norms as mental objects, i.e. "hybrid configurations of beliefs and goals", as Conte and Castelfranchi do (1995: 192), then there is no obvious reason to assume that behaviorally enforced goals and beliefs do not in turn promote habitual and at the same time norm-abiding behavior.

What, then, is the crucial role of emotion in these processes? It has been repeatedly shown that the mechanisms of association and confirmation, herein proposed to enforce goals and beliefs, are inextricably tied to cognitive-emotional interactions, for example in memory formation, stimulus-reinforcement learning, mood congruency, state-dependent recall of information, and social judgments (Forgas, 1995, 2000; Clore, Schwarz, and Conway, 1994; Rolls, 2004; Bless, 2000). From a design perspective, in layered architectures emotion ultimately facilitates the processes constituting the supplementary automatic, associative layer on top of the procedural reasoning and reactive layers, for example in Sloman's (2001) "CogAff" system, Staller and Petta's (2001) TABASCO architecture, in Castelfranchi's hybrid model (Castelfranchi, 2000, 2001) and in our own work (von Scheve and Moldt, 2004). This perspective is the outcome of diverse empirical and theoretical investigations into the inextricability of cognition and emotion, which we cannot account for in its entirety here (cf. Power and Dalgleish, 1997; Frijda, 1994; LeDoux, 1994). Thus, even if we conceive of normative behavior as, and only as, conscious intentional behavior, there is no reasonable way of bypassing emotion as long as one is committed to modeling socially intelligent behavior in human and non-human animals. Jon Elster has formulated this position to the point in paraphrasing Max Weber, holding that "a social norm is not like a taxi from which one can disembark at will" (Elster, 1989: 106). Also, regarding the issues outlined by Brooks (1991), a behavioral approach to norms might at least be an alternative worth exploring in multi-agent systems research.

Emotions enforce and maintain social norms

Besides the general influence of emotion on unintentional normative and socially functional behavior just portrayed, there is another noteworthy social function of emotion strongly related to social norms that is particularly evident in social interactions. This relation not only points to the fact that normative behavior is guided by emotions, but also to the fact that the maintenance and enforcement of norms is closely tied to emotion as well. The further analysis of this perspective strongly suggests an approach also taken by Conte and Castelfranchi (1995), namely to view a norm—as such—as a mental object, i.e. a hybrid configuration of beliefs and goals (which in no way, as illustrated above, underestimates the power of unintentional normative behavior). Endorsing this approach, the social, temporal, and spatial distribution of norms makes them instances of a macro system; at the same time, however, their definition as configurations of propositional attitudes (i.e. beliefs and goals) renders them an instance of the micro level and accordingly a primary subject matter of different kinds of reasoning processes (Fodor, 1981; Dretske, 1988). Conte and Castelfranchi (1995) argue in analogy, defining norms as directives or instructions which are represented as beliefs, substantially determining behavior by generating new goals, i.e. normative goals. 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) remain largely unanswered, though a satisfactory answer is essential if one wants to find a solution for the "foundational theoretical problem" mentioned by Castelfranchi.

In what follows, we argue that emotion might be a loophole here and that Jon Elster's (1999) conceptualization thereof is suited to serve as an addition to existing positions. Elster delivers a definition of certain qualities of social norms rather than of the concept of a social norm itself. Accordingly, social norms can be described as follows (Elster, 1999: 145f): First, social norms are non-outcome-oriented phenomena. They can have unconditional imperative character but also conditional if they refer to past actions. Second, social norms are shared with other members of a social system in which the process of sharing is also socially shared. Third, and this is a consequence of the second characteristic, norms are subject to direct and third-party punishment, whereas punishing is prescribed by a norm.

Some evidence

When analyzing the question why agents adhere to norms and social systems exhibit large-scale normative behavior, it is critical to examine the kinds of sanctions in case of non-compliance with norms. Particularly in economic theory sanctions resulting from non-compliance are described as a withdrawal of material resources (Becker, 1976). Withdrawing or withholding material resources (e.g. loss of money or other resources, refusal to cooperate, breach of a contract), however, is by no means the definitive or most efficient way to successful sanctioning. Even more decisive in this respect is the fact that norm-violators usually interpret material sanctions also as a vehicle for the expression of strong negative emotions, e.g. contempt, disdain, detestation, or disgust. In consequence, violators feel shame or guilt. The experience of shame in this respect will be even worse because—and in contrast to guilt—the perspective of the sanctioner is incorporated and accounted for more vividly. Furthermore, it has been argued that shame indicates severe threat to an actor's social bonds (Scheff, 2003).

In this context, Jon Elster has shown that the *material* aspect of sanctions lies solely within the question of how much it costs the *punisher* to impose the sanction, and not within 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 an intended sanction, the more vividly aware is the offender of the negative emotions accompanying the implementation of these sanctions, and the more strongly the offender will feel the resulting shame. The sanctioning costs a punisher is willing to accept signals to the offender the severity of the norm digression. It is no exception to the rule that punishers accept enormous costs by far outreaching the damage an offender has caused. But these costs are by no means futile, quite the contrary they exemplify the negative emotional meaning that accompanies the sanction and emphasize the offender's obligation to feel guilt or shame.

These mechanisms have also been validated empirically and experimentally. For example in one-shot ultimatum games, participants carry out altruistic punishment at their own costs with no material rewards or possible deterrence, thereby enforcing the norm of fairness. Ernst Fehr and his group (Fehr and Gaechter, 2002) have shown that emotion is a proximate cause of altruistic punishment. Their experiment illustrates that free riding in public goods games, i.e. the violation of the norm of reciprocal cooperation, reliably causes strong negative emotions, that most free riders do indeed expect these emotions, and that they in turn trigger punishment. Fehr and Gaechter conclude that cooperation thrives if altruistic punishment is possible, and collapses if altruistic punishment is ruled out. In another experiment Sanfey and colleagues (2003) have shown that unfair offers in an ultimatum game, i.e. the violation of fairness norms, reliably modulate anterior cingulate activity, an area of the human brain that is "responsible" for negative affects such as pain and distress, thirst and hunger. Another astonishing study has shown that punishment of norm violators activates human's neural reward-circuitry (mainly the dorsal striatum), indicating that actors derive satisfaction from punishing defectors (de Quervain et al., 2004). A well known side effect in this case usually is the relief experienced from one's anger ceasing as a consequence of delivered punishment.

Further evidence comes from studies analyzing the violator's perspective. Naomi Eisenberger and colleagues found that social rejection, i.e. social exclusion or even ostracism, possibly as a consequence of the violation of a social norm, modulates human's anterior cingulate cortex, an area usually "responsible" for processing physical pain. They conclude that social pain, as experienced in shame or guilt states, and physical pain share some of the same neural circuits and corresponding computational mechanisms (Eisenberger, Lieberman, and Williams, 2003; Eisenberger and Lieberman, 2004). Still others found that the perception of a norm transgression activates specific brain regions usually involved in representing aversive emotional reactions in others (Berthoz et al., 2002).

Action incentives

The evidence just outlined to us suggests that norms may well be represented as propositional attitudes, but in order to function as such, i.e. to generate normative behavior and to be enforced and maintained, emotion is needed. What makes actors responsive to social norms is neither external punishment nor rational interest alone. It is (also) the anticipation of negative emotions that arise in case of a norm violation and the satisfaction derived from punishing violators. Now, what consequences does “emotional deterrence” have for one’s options to act? Striving for emotional gratification, i.e. the motivation to seek out interactions resulting in positive emotions and to avoid those resulting in negative emotions is considered a basic motivator of human behavior (Turner, 1999; Collins, 1984). Thus, emotions can well relate to future actions by substantially affecting the generation of plans and many other cognitive competencies (Loewenstein and Lerner, 2003). Anthony Giddens, for example, considers concerns about the loss of ontological security as a central aspect in his theoretical framework: fear of loosing ontological security and facticity is a primary motivator of social action (Giddens, 1991). Other authors, for example Randall Collins (1984) or Michael Hammond (1991), who regard emotional gratification as a motivator of action directly scalable to large-scale contexts, assume that actors have an inborn need for positive emotional exchange processes, which may solidify to “interaction ritual chains” and thus contribute to the emergence of robust social structures (Collins, 1981; Collins, 2004).

In view of the interaction of emotion, norms, and normative behavior, we can further assume that shame and contempt in particular serve as a vehicle for maintaining norms by generating normative goals 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 resources through sanctions, but instead as a result of the fear of emotion-driven sanctions (by way of negative emotions such as contempt, disdain, detestation or disgust in the punisher), that would result in negative emotions, e.g., shame, guilt, or embarrassment, in the violator.

An approach to modeling

Improving AI systems by taking into account emotion or functionally equivalent mechanisms is not a new idea, its origins rooted in the contributions of researchers like Herbert Simon (1967), Aaron Sloman and Monica Croucher (1981) or Marvin Minsky (1986). In the late 1980s first reviews of AI-models of emotion were assembled (Dyer, 1987; Pfeifer, 1988; Ruebenstrunk, 1998) and until now, research on emotion within computer science has mainly revealed three basic motivations: performance, human-computer interaction, and the simulation of naturally occurring phenomena (Picard, 1997; Wehrle, 1998; Scheutz, 2002). As indicated above, research on emotional agents has largely been concerned either with isolated entities or dyadic interaction settings (agent-agent and agent-user) and almost all of the models in question are based on psychological and neuroscientific theories of emotion, as we have delineated for the field of human-computer interaction previously (Moldt and von Scheve, 2002, 2002a). However, as we have briefly shown in the preceding sections, emotion is highly socially functional not only in dyadic settings but also in large-scale social systems, one of the facilitators of functionality being the interplay with social norms. Therefore it seems evident that if norms are employed as a coordination principle in multi-agent systems, the use of emotion suggests itself.

Our approach to modeling the interdependencies and reciprocities of the social functions of emotion first and foremost aims at locating emotion in a micro-macro framework, as initially outlined above. This is further facilitated by taking a layered perspective on sociality that can be represented by the multi-agent architecture MULAN, providing a conceptually highly flexible framework. The MULAN architecture is a technical implementation of the agent concept that may simultaneously account for almost any aspect of an application model. Separation of the two perspectives is achieved by the SONAR agent architecture in making use of an array of sociological concepts (cf. von Luede, Moldt, and Valk, 2003; Koehler et al., 2005). In particular, for each and every social entity constituting the different societal layers—i.e. actors, social processes, and social structures—a single SONAR agent is deployed and made available on the MULAN platform. The social units' inherent logics are in turn represented by multi-agent systems which are directly subordinated to the respective units. This framework has proven to be quite useful in modeling micro-macro aspects of behavior in organizational and institutional social contexts (cf. von Luede, Moldt, and Valk, 2003).

Representing emotion and norms in SONAR

Michael Koehler and colleagues (2003) provide a modeling framework conceiving of social entities, i.e. actors, processes, and structures, as first-order objects that can be modeled side by side simultaneously. The framework allows the thorough representation of direct interdependencies occurring on identical layers of observation (abstraction), whereas the internal properties of a particular social entity may still be entirely encapsulated from direct access by other social entities of the same layer. The social entities' internal logics are largely autonomous and may thus differ from one another substantially. For example, any actor might maintain an arbitrarily complex representation of its (social) environment, which is in turn represented as a set of specifically networked social entities, i.e. SONAR-based multi-agent systems. The system can be quite simple in case of some primitive agents, it can become, however, highly complex when agents are characterized by increasing degrees of social differentiation.

Imagine, for example, the mind of some muddle-headed professor and her representations of the external and internal world in its entire complexity—including all mental contents and inconsistencies. Being an exemplification, few would claim to be able to model a real human being's mental contents. Rather, the designer might well simplify or extend the model according to the requirements of the task at hand. The same would also be true for the social processes that are usually established between social agents as well as for considerably stable (but not unalterable) social structures. Therefore, social entities in this framework *in themselves* contain the necessary references to other entities, or, to be more explicit, are to some extent *made up of* these references. Evidently, these references do not have to be modeled independently. Part of this architectural layout indeed reflects what human actors do in everyday social interactions: continuously modeling (representing) other actors and mentally relating to them, thereby constituting what is dubbed in the social sciences as "intersubjectivity". And, even more important, the structural layer, i.e. social structural effects, to some extent determine the references and representations on the lower layers while in turn being constituted by them through, e.g., shared knowledge, shared beliefs, and collective acceptance (Tuomela, 2002; Searle, 1995).

In designing a specific system, the designer is required to deliberately select specific aspects of a system which are supposed to be crucial for the task at hand. SONAR models can then exactly specify the model's most important elements. Whereas the MULAN environment provides the technical framework for implementing key concepts like autonomy, mobility, cooperation, and adaptation, the SONAR architecture is used to model the *internal* representations of an entity by means of a multi-agent system. Figure 1 illustrates the interplay of two autonomous social entities, in this case an actor and the social process it is involved in. Social structural embedding is achieved by the social processes that are involved in the evolution, reinforcement, and reproduction of new structures. Conceptually, we have to deal with the same kind of connection between both parts of the model by means of synchronous channels (see below).

Figure 1 depicts a powerful variant of Petri nets—a reference net (Valk, 1998). Rectangles (transitions) represent actions, whereas circles (places) denote available or unavailable resources or conditions that may be fulfilled. An arc determines the specific context of the transition. Thus, arcs that are directed from transitions to places can be interpreted as preconditions for action, whereas arcs that are directed from places to transitions represent actions' outcomes. A firing transition (or an action that is implemented) will remove resources or conditions (for short: tokens) from places and insert them into some other place. A notable property of a reference net is the possibility that the tokens located on a place in a net (the *system net*) may again be a reference net (an *object net*) (or any other Java object) (Valk, 1998). Object net and system net are synchronized through synchronous channels, whereby one of the nets is expected to *reference* the other one. For example, as illustrated in figure one, actions of an actor are synchronized with some social process by two transitions constituting the synchronous channel: "observe" and "act" in the actor net, in this case modeled as an object net, and "access observability" and "expand action repertoire" in the social process net, in this case the actor net's system net. Needless to say, tokens $a1$ and $a2$ could also be modeled as additional actor (object) nets.

We should point out that an increase in resolution is facilitated by "zooming" into the references to the relevant actor nets. As we have said before, any social entity is in itself essentially autonomous. In principle, the mutual relationships between them are assumed to be symmetrical, even though modeling them might imply some (rational) hierarchical structure. The social interactional concepts we have used, e.g. "observation" and "action", are accounted for solely in the process entity. Any adjustment

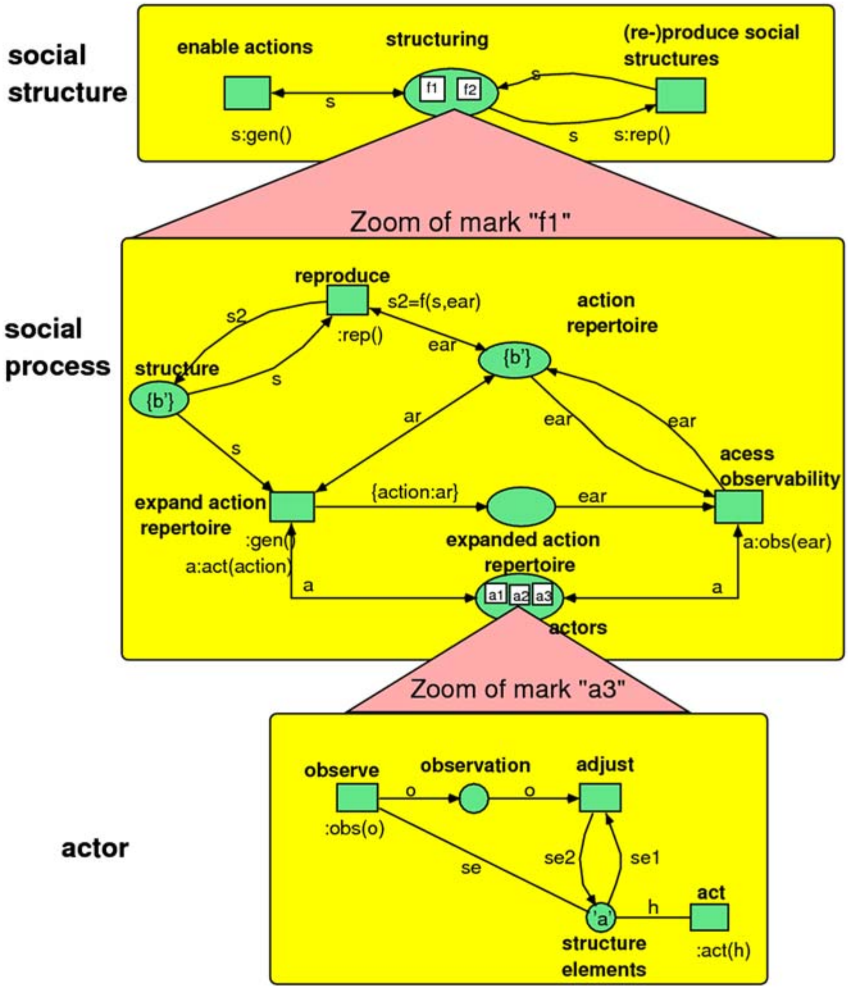

Fig. 1 A reference net representing actor, social process, and social structure

succeeds locally, without direct interaction with the social environment. Further details had to be omitted here, but can be found in either von Luede, Moldt, and Valk (2003) or in Koehler et al. (2005).

Although we have already used a range of concepts from the social sciences within the SONAR models, for example acknowledgement, observation, action, etc., the role of emotion has hardly been taken into account. Hence, we propose some simple but fundamental enhancements to SONAR that still allow for a conventional approach to modeling, therefore treating all aspects concerning emotion separately for the time being. This requires some explicit decisions in view of what belongs to an emotion proper and what does not. Given this perspective, the problem of adequately connecting both models becomes quite obvious. We aim at solving this issue in a simple and homogenous manner with the applied modeling techniques: existing MULAN models of organizational micro-macro behavior are to be completed with additional models of the functions of emotion in large-scale systems.

Corresponding findings in emotion theory rather explicitly dealing with the nature of emotion as such might be represented in a segregated emotion model. On the other hand, in case the emotion theories under scrutiny state explicit references to social processes and structures (e.g., sociological models), they may well be linked to the conventional, rational elements of the model by synchronous channels, thus creating a kind of homologous "parallel world". Modeling each and every social entity as an individual net allows depicting their mutual relations by means of references in the sense of a "pointer". Synchronous channels in this case may represent any possible interconnection of system behavior in distinct parts of the model.

We are well aware of the fact that the explicit separation of emotion from other parts of the model is prone to critique in that it seemingly promotes a dichotomy between "rational" and "emotional" perspectives that most researchers would nowadays consider obsolete. However, we favor an initial explicit technical separation and simplification in order to achieve just that: to be able to illustrate precisely and concisely in semi-formal semantics the mutual and supposedly reciprocal and recursive interconnections between the two domains. The necessity for such models in social science research has been put forward by Jonathan Turner: "Such models [...] can provide a picture of process—that is, of how variables influence each other across time. Moreover, they can also give us a view of complex causal processes. Too often in sociology, we employ simple causal models [...] that document one-way causal chains among empirical indicators of independent, intervening, and dependent variables. But actual social processes are much more complex, involving feedback loops, reciprocal causal effects, lag effects, threshold effects, and the like" (Turner, 1988: 17).

Obviously, this is no means to overcome general system complexity, although complexity can be allocated to different layers, their integration demanding still further efforts because the linkages between the different layers are to be located and modeled explicitly. (In this respect we might, for example, apply solutions of similar problems found in the combination of different views in models created with the unified modeling language.) Although at some point an integration of the two perspectives is necessary and desired, we currently favor the advantages of an explicit separation, in particular the intuitive simplicity of the detached points of view.

Smaller and less complex models can usually be implemented much faster with programming languages that do not distinguish between different points of view. However, the separation of different points of view in order to handle a system's complexity is a common method in computer science. Apart from issues of technical complexity, emotion (and other) theories in the social and behavioral sciences are also highly complex systems that might demand even more flexibility as is generally required in construction-oriented computer science models. Therefore, a separation into different layers seems to be useful not only for social theory but also in view of a sharper examination of different analytical layers. The possibility of modeling discrete emotions and their specific components as explicit states and processes which are integrated into the existing (rational) models still remains largely unelaborated, albeit it is one goal of our future work.

Discussion

Finally, we will very briefly evaluate the modeling approach presented above as well as the possibilities it holds in view of modeling the interrelation of norms and emotion. Generally speaking, accounting for emotion in our modeling and simulation framework of behavior in organizational and institutional social contexts incorporates further aspects of reality that might contribute to better solutions, particularly when it comes to modeling informal social interactions. In addition, modeling the social functions of emotion is supposed to improve multi-agent systems and normative systems in particular, for example in view of alternative coordination solutions. However, strong evidence has yet to be presented that AI-models of emotion do indeed facilitate better solutions.

In addition, the approach allows the investigation of the relationship between cognition and emotion by means of explicitly separated components. This is particularly important in view of the interaction of norms as explicitly represented mental objects and emotions as processes with non-propositional output. The analysis of different components of the emotion process belonging to different representational formats is also of interest for original emotion research. Currently, there is much debate on the question of how information that is represented in different representational formats and memory structures (i.e., semantic vs. non-propositional) interacts in the generation of emotion (cf. Reisenzein, 2001).

Thus, the interrelation between social norms and emotion—and between two different representational systems—can be examined on different levels of social abstraction. On the structural layer, norms may either be related to concepts largely disregarding emotion, or with those considering emotion on all layers of a model (actor, process, and structure). This openness and flexibility permits an analysis of different theories and also computational models of emotion *simultaneously* and to different extents. One of the core advantages of the framework is that existing emotion-based agent architectures which are primarily concerned with representing actor- and eventually process-layers (e.g., Staller and Petta, 1998, Sloman, 2001; Cami, Lisetti, and Sierhuis, 2004)) might well be integrated (and extended), probably on the protocol-layer, as emotion generating entities.

References

Abe JA, Izard CE (1999) The developmental functions of emotions: An analysis in terms of differential emotions theory. Cog Emotion 13(5):523–549

de Almeida LB, da Silva BC, Bazzan AL (2004) Towards a physiological model of emotions: First steps. In: Proceedings of the 2004 AAAI Spring Symposium. AAAI Press, Menlo Park, CA

Averill JR (1992) The structural bases of emotional behavior. In: Clark MS (ed) Emotion. Sage, Newbury Park, CA, pp 1–24

Avila-Garcia O, Canamero L (2005) Hormonal modulation of perception in motivation-based action selection architectures. In: Proceedings of Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action. Hatfield, UK

Aube M, Senteni A (1996) Emotions as commitments operators: A foundation for control structure in multi-agents systems. In: Proceedings of the 7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'96). LNAI Vol. 1038. Springer, Berlin, pp 13–25

Axelrod R (1986) An evolutionary approach to norms. Am Political Sci Rev 80(4):1095–1111
Bazzan AL, Adamatti DF, Bordini R (2002) Extending the computational study of social norms with a systematic model of emotions. In: Proceedings of the 16th Brazilian Symposium on Artificial Intelligence (SBIA'02). LNCS Vol. 2507. Springer, Berlin, pp 108–117
Bendor J, Swistak P (2001) The evolution of norms. Am J Sociol 106(6):1493–1545
Becker G (1976) The economic approach to human behavior. University of Chicago Press, Chicago, IL
Berthoz S, Armony JL, Blair RJ, Dolan RJ (2002) An fMRI study of intentional and unintentional (embarrassing) violations of social norms. Brain 125(8):1696–1708
Bless H (2000) The interplay of affect and cognition: The mediating role of general knowledge structures. In: Forgas JP (ed) Feeling and Thinking. Cambridge University Press, New York, NY, pp 201–222
Brooks RA (1991) Intelligence without representation. Artif Intell 47(1–3), 139–159
Cacioppo JT, Berntson GG, Larsen JT, Poehlmann KM, Ito TA (2000) The psychophysiology of emotion. In: Lewis R, Haviland-Jones JM (eds) Handbook of emotions, 2nd Ed. Guilford, New York, NY pp 173–191
Cami A, Lisetti C, Sierhuis M (2004) Towards the simulation of a multi-level model of human emotions. In: Proceedings of the 2004 AAAI Spring Symposium. AAAI Press, Menlo Park, CA
Canamero D (ed) (1998) Emotional and intelligent: The tangled knot of cognition. In: Proceedings of the 1998 AAAI Fall Symposium. Menlo Park, CA: AAAI Press
Carley K (1986) An approach for relating social structure to cognitive structure. J Math Sociol 12(2):137–189
Castelfranchi C (2000) Affective appraisal vs. cognitive evaluation in social emotions and interactions. In: Paiva A (ed) Affect in interactions. Springer, Heidelberg
Castelfranchi C (2001) The theory of social functions: Challenges for computational social science and multi-agent learning. Cogn Sys Res 2(1):5–38
Castelfranchi C, Dignum F, Jonker CM, Treur J (1999) Deliberative normative agents: Principles and architecture. In: Proceedings of the 6th International Workshop on Intelligent Agents VI, Agent Theories, Architectures, and Languages (ATAL). LNCS Vol. 1757. Springer, London, pp 364–378
Clarke S (2003) Psychoanalytic sociology and the interpretation of emotion. J Theory Soc Behav 33(2):143–161
Clore GL, Schwarz N, Conway M (1994) Affective causes and consequences of social information processing. In: Wyer RS, Srull TK (eds) Handbook of social cognition. 2nd Ed. Erlbaum, Hillsdale, NJ, pp 323–417
Clore GL (1994) Why emotions are felt. In: Ekman P, Davidson RJ (eds) The nature of emotion. Oxford University Press, New York, NY pp 103–111
Coleman JS (1990) Foundations of social theory. Harvard University Press, Cambridge, MA
Collins R (1981) On the microfoundations of macrosociology. Amer J Sociol 86(5):984–1014
Collins R (1984) The role of emotion in social structure. In: Scherer KR, Ekman P (eds) Approaches to Emotion. Erlbaum, Hillsdale, NJ, pp 385–396
Collins R (2004) Interaction ritual chains. Princeton University Press, Princeton, NJ
Conte R, Castelfranchi C (1995) Norms as mental objects. In: Proceedings of the 5th European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW '93). LNAI Vol. 957. Springer, Berlin, pp 186–196
Cosmides L, Tooby J (2000) Evolutionary psychology and the emotions. In: Lewis R, Haviland-Jones JM (eds) Handbook of emotions, 2nd Ed. Guilford, New York, NY pp 91–115
Dignum F (1999) Autonomous agents with norms. Artif Intell Law 7(1):69–79
Dretske F (1988) Explaining behavior. MIT Press, Cambridge, MA
Dyer MG (1987) Emotions and their computations: Three computer models. Cogn Emotion 1(3):323–347
Eisenberger NI, Lieberman MD, Williams KD (2003) Does rejection hurt? An fMRI study of social exclusion. Sci 302:290–292
Eisenberger NI, Lieberman MD (2004) Why rejection hurts: A common neural alarm system for physical and social pain. Trends Cogn Sci 8(7):294–300
Elliot CD (1993) Using the affective reasoner to support social simulations. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI'93). Morgan Kaufmann, Los Altos, CA, pp 194–200
Elliot CD (1992) The affective reasoner. A process model of emotions in a multi-agent system. PhD Thesis, Intitute for the Learning Sciences, Northwestern University
Elster J (1989) Social norms and economic theory. J Econ Perspect 3(4):99–117

Elster J (1996) Rationality and the emotions. The Econ J 106(438):1386–1397
Elster J (1999) Alchemies of the mind. Cambridge University Press, Cambridge
Elster J (2004) Emotions and rationality. In: Manstead AS, Frijda NH, Fischer A (eds) Feelings and emotions. Oxford University Press, New York, NY, pp 30–48
Engel P (2002) Intentionality, normativity and community. Facta Philosophica 4(1):25–49
Fehr E, Fischbacher U (2004) Social norms and human cooperation. Trends Cogn Sci 8(4):185–190
Fehr E, Fischbacher U (2004a) Third-party punishment and social norms. Evol Hum Behav 25(1):63–87
Fehr E, Gaechter S (2002) Altruistic punishment in humans. Nature 415: 137–140
Fix J (2004) Emotionale agenten. Diploma Thesis. University of Hamburg, Computer Science Department
Fodor J (1981) Propositional attitudes. In: Fodor J (ed) RePresentations. MIT Press, Cambridge, MA, pp 177–203
Forgas JP (1995) Mood and judgment: The affect infusion model (AIM). Psychological Bulletin 117(1):39–66
Forgas JP (2000) Feeling is believing? The role of processing strategies in mediating affective influences on beliefs. In: Frijda NH, Manstead A, Bem S (eds) Emotions and beliefs. University Press Cambridge, Cambridge, pp 108–143
Frank RH (1993) The strategic role of emotions: Reconciling over- and undersocialized accounts of behaviour. Rationality and Society 5(2):160–184
Frank RH (2001) Cooperation through emotional commitment. In: Nesse RM (ed) Evolution and the capacity for commitment. Russel Sage, New York, pp 57–76
Frijda NH (1986) The emotions. Cambridge University Press, Cambridge
Frijda NH (1994) Emotions require cognitions, even if simple ones. In: Ekman P, Davidson RJ (eds) The nature of emotion. Oxford University Press, New York, NY, pp 197–202
Gerhards J (1988) Soziologie der emotionen. Juventa, Weinheim
Giddens A (1991) Modernity and self-identity. Polity Press, Cambridge
Gmytrasiewicz PJ, Lisetti CL (2000) Using decision theory to formalize emotions for multi-agent systems. In: Proceedings of the 2nd ICMAS'00 Workshop on Game Theoretic and Decision Theoretic Agents. Boston, MA
Hatano G, Okada N, Tanabe H (eds) (2000) Affective minds. In: Proceedings of the 13th Toyota Conference. Elsevier, Amsterdam
Heckathorn DD (1989) Cognitive science, sociology, and the theoretic analysis of complex systems. J Math Sociol 14(2/3):97–110
Heise DR, Calhan C (1995) Emotion norms in interpersonal events. Soc Psychol Q 58(4):223–240
Heise D, O'Brien J (1993) Emotion expression in groups. In: Lewis M, Haviland JM (eds) Handbook of emotions. Guilford, New York, NY, pp 489–497
Hochschild AR (1979) Emotion work, feeling rules, and social structure. Ame J Sociol 85(3):551–575
Horne C (2001) Sociological perspectives on the emergence of norms. In: Hechter M, Opp K-D (eds) Social norms. Russell Sage, New York, NY, pp 3–34
Horne C, Cutlip A (2002) Sanctioning costs and norm enforcement. Rationality and Society 14(3):285–307
Jennings NR (2000) On agent-based software engineering. Artif Intell 117(2):277–296
Keltner D, Gross JJ (1999) Functional accounts of emotion. Cogn Emotion 13(5):467–480
Keltner D, Haidt J (1999) Social functions of emotion at four levels of analysis. Cogn Emotion 13(5):505–521
Kemper TD (1978) A social interactional theory of emotions. Wiley, New York, NY
Knorr-Cetina KD (1981) The micro-sociological challenge of macro-sociology. In: Knorr-Cetina KD, Cicourel AV (eds) Advances in social theory and methodology. Routledge & Kegan Paul, Boston, MA, pp 1–47
Koehler M, Moldt D, Roelke H (2001) Modeling the structure and behaviour of petri net agents. In: Proceedings of the 22nd International Conference on Application and Theory of Petri Nets (ICATPN'01). LNCS Vol. 2075. Springer, Berlin, pp 224–241
Koehler M, Moldt D, Roelke H (2003) 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 (RASTA'03). Edinburgh

Koehler M, Moldt D, Roelke H, Valk R (2005) Linking micro and macro descriptions of scalable social systems using reference nets. In: Fischer K, Florian M (eds) Socionics: Its contribution to the scalability of complex social systems. LNCS. Springer, Heidelberg [in press]

Lawler EJ, Ridgeway C, Markowsky B (1993) Structural social psychology and the micro-macro problem. Sociol Theory 11(3):268–290

Lawler EJ, Thye SR, Yoon J (2000) Emotion and group cohesion in productive exchange. Ame J Sociol 106(3):616–657

LeDoux JE (1996) Cognitive emotional interactions in the brain. In: Ekman P, Davidson RJ (eds) The nature of emotion. Oxford University Press, New York, NY, pp 216–223

Lizardo O (2004) The cognitive origins of bourdieu's habitus. Journal for the Theory Soc Behav 34(4):375–401

Loewenstein G, Lerner JS (2003) The role of affect in decision making. In: Davidson RJ, Scherer KR, Goldsmith HH (eds) Handbook of affective sciences. Oxford University Press, New York, NY, pp 619–642

von Luede R, Moldt D, Valk R (eds) (2003) Sozionik: Modellierung soziologischer Theorie. [Socionics: Modeling Sociological Theory]. Muenster: Lit

Macy MW, Willer R (2002) From factors to actors: Computational sociology and agent-based modeling. Annu Rev Sociol 28:143–166

Malsch T (2001) Naming the unnamable: Socionics or the sociological turn of/to distributed artificial intelligence. Autonomous Agents and Multi-Agent Systems 4(3):155–186

Mendao M (2004) Hormonally moderated neural control. In: Proceedings of the 2004 AAAI Spring Symposium. AAAI Press, Menlo Park, CA

Moldt D, von Scheve C (2002) Emotions in hybrid social aggregates. In: Herczeg M, Prinz W, Oberquelle H (eds) Mensch & Computer 2002. Teubner, Stuttgart, pp 343–352

Moldt D, von Scheve C (2002a) Attribution and adaptation: The case of social norms and emotion in human-agent interaction. In: Proceedings of The Philosophy and Design of Socially Adept Technologies. Workshop held in conjunction with CHI'02. Minneapolis, MN

Minsky M (1986) The society of mind. Simon & Schuster

Oatley K, Jenkins JM (1996) Understanding emotions. Basil Blackwell, Oxford

Opp K-D (2002) When do norms emerge by human design and when by the unintended consequences of human action? The example of the no-smoking norm. Rationality and Society 14(2):131–158

Panzarasa P, Jennings NR (2002) Social influence, negotiation and cognition. Simulation Modelling Practice and Theory 10(5–7):417–453

Panzarasa P, Norman TJ, Jennings NR (2001) Social mental shaping: Modelling the impact of sociality on autonomous agents' mental states. Comput Intell 17(4):738–782

Pfeifer R (1998) Artificial intelligence models of emotion. In: Hamilton V, Bower GH, Frijda NH (eds) Cognitive perspectives on emotion and motivation. Kluwer, Dordrecht, pp 287–320

Picard RW (1997) Affective computing. MIT Press, Cambridge, MA

Power M, Dalgleish T (1997) Cognition and emotion: From Ordeo to disorder. Psychology Press, Hove

de Quervain DJ-F, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, Fehr E (2004) The neural basis of altruistic punishment. Sci 305:1254–1258

Reisenzein R (2001) Appraisal processes conceptualized from a schema theoretic perspective. In: Scherer KR, Schorr A, Johnstone T (eds) Appraisal processes in emotion. Oxford University Press, New York, NY, pp 187–204

Rolls ET (2004) The functions of the orbitofrontal cortex. Brain and Cognition 55(1):11–29

Ruebenstrunk G (1998) Emotional computers. Computer models of emotions and their meaning for emotion-psychological research. Unpublsihed Manuscript, available from http://www.schooldays.de/ruebentemp/emeocomp/content.HTM

Saam NJ, Harrer A (1999) Simulating norms, social inequality, and functional change in artificial societies. Journal of Artificial Societies and Social Simulation, 2(1). http://www.soc.surrey.ac.uk/JASSS/2/1/2.html

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the ultimatum game. Sci 300:1755–1758

Sawyer RK (2003) Artificial societies. Multiagent systems and the micro-macro link in sociological theory. Sociol Meth & Res 31(3):325–363

Sawyer RK (2004) Social explanation and computational simulation. Philosophical Explorations 7(3):219–231

Scheff TJ (2003) Shame in self and society. Symbolic Interaction 26(2):239–262
Scheutz M (2002) Agents with or without Emotions? In: Proceedings of the 15th International Florida Artificial Intelligence Symposium (FLAIRS'02). AAAI Press, Menlo Park, CA, pp 89–94
von Scheve C, Moldt D (2004) Emotion: Theoretical investigations and implications for artificial social aggregates. In: Lindeman G, Moldt D, Paolucci P (eds) Regulated agent-based social systems. LNAI Vol. 2934. Springer, Berlin, pp 189–209
von Scheve C, von Luede R (2005) Emotion and social structures. Towards an interdisciplinary approach. J Theory Soc Behav 35(3):303–328
Schwarz N, Clore GL (1988) How do i feel about it? The informative function of affective states. In: Fiedler K, Forgas JP (eds) Affect, cognition, and social behavior. Hogrefe, Toronto, pp 44–62
Searle JR (1995) The construction of social reality. Free Press, New York, NY
Shoham Y, Tennenholtz M (1997) On the emergence of social conventions: Modeling, analysis, and simulations. Artifl Intell 94(1–2):139–166
Simon HA (1967) Motivational and emotional controls of cognition. Psychological Review 74:29–39
Sloman A (2001) Varieties of affect and the cogaff architecture schema. In: Proceedings of the AISB'01 Symposium on Emotion, Cognition, and Affective Computing. SSAISB, York, pp 39–48
Sloman A, Croucher M (1981) Why robots will have emotions. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI'81). Vancouver, pp 197–202
Smith CA, Pope LK (1992) Appraisal and emotion. The interactional contributions of dispositional and situational factors. In: Clark MS (ed) Emotion and social behavior. Sage, Newbury Park, CA, pp 32–62
Staller A, Petta P (1998) Towards a tractable appraisal-based architecture for situated cognizers. In: Proceedings of Grounding Emotions in Adaptive Systems. Workshop Notes of the 5th International Conference of the Society for Adaptive Behaviour (SAB'98). Zurich, pp 56–61
Staller A, Petta P (2001) Introducing emotions into the computational study of social norms: A first evaluation. Journal of Artificial Societies and Social Simulation 4(1). http://soc.surrey.ac.uk/JASSS/4/1/2.html
Trappl R, Petta P, Payr S (eds) (2003) Emotions in humans and artefacts. MIT Press, Cambridge, MA
Tuomela R (1995) Norms and agreement. European Journal of Law, Philosophy and Computer Science 5(1):41–46
Tuomela R (2003) Collective acceptance, social institutions, and social reality. Am J Econ Sociol 62(1):123–165
Turner JH (1988) A theory of social interaction. Stanford University Press, Stanford, CA
Turner JH (1999) Toward a general sociological theory of emotions. Journal for the Theory of Social Behaviour 29(2):133–161
Valk R (1998) Petri nets as token objects. An introduction to elementary object nets. In: Proceedings of Application and Theory of Petri Nets. Lisbon, Portugal. LNCS Vol. 1420. Springer, Berlin, pp 1–25
Verhagen H (2000) Norm autonomous agents. PhD Thesis, Royal Institute of Technology and Stockholm University
Wehrle T (1998) Motivations behind modeling emotional agents: Whose emotions does your robot have? In: Proceedings of Grounding Emotions in Adaptive Systems. Workshop Notes of the 5th International Conference of the Society for Adaptive Behaviour (SAB'98). Zurich
Wiley N (1988) The micro-macro problem in social theory. Sociol Theory 6(2):254–261
Wooldridge M (2000) Reasoning about rational agents. MIT Press, Cambridge, MA

Christian von Scheve graduated in Sociology with minors in Psychology, Economics, and Political Science at the University of Hamburg, where he also worked as a research assistant at the Institute of Sociology. Currently, he is a 3rd year PhD student at the University of Hamburg. He was a Fellow of the Research Group "Emotions as Bio-Cultural Processes" at the Center for Interdisciplinary Research (ZiF) at Bielefeld University. In his doctoral thesis he develops an interdisciplinary approach to emotion and social structural dynamics, integrating emotion theories from the neurosciences, psychology, and the social sciences. He has published on the role of emotion in large-scale social systems, human-computer interaction, and multi-agent systems. He is co-editor of a forthcoming volume on emotion regulation.

Daniel Moldt received his BSc in Computer Science/Software Engineering from the University of Birmingham (England) in 1984, graduated in Informatics at the University of Hamburg, with a minor in Economics in 1990. He received his PhD in Informatics from the University of Hamburg in 1996, where he has been a researcher and lecturer at the Department of Informatics since 1990. Daniel Moldt is also the head of the Laboratory for Agent-Oriented Systems (LAOS) of the theoretical foundations group at the Department of Informatics. His research interests focus on theoretical foundations, software engineering and distributed systems with an emphasis on agent technology, Petri nets, specification languages, intra- and inter-organizational application development, Socionics and emotion in informatics.

Julia Fix is currently a PhD student at the Theoretical Foundations of Computer Science Group, Department for Informatics at the University of Hamburg. She studied Informatics and Psychology at the University of Hamburg, with an emphasis on theoretical foundations of multi-agent systems and wrote her diploma theses about emotional agent systems. Her current research interests focus on conceptual challenges and theoretical foundations of modelling emotions in multi-agent systems, emotion-based norm enforcement and maintenance, and Socionics. A further research focus are Petri nets, in particular the use of Petri-net modelling formalisms for representing different aspects of emotion in agent systems.

Rolf von Lüde is a professor of Sociology at the University of Hamburg with a focus in teaching and research in Sociology of Organizations, Work and Industry since 1996. He graduated in Economics, Sociology, and Psychology, and received his doctorate in Economics and the venia legendi in Sociology from the University of Dortmund. His current research focuses on labor conditions, the organization of production, social change and the educational system, the organizational structures of university, Socionics as a new approach to distributed artificial intelligence in cooperation with computer scientists, new public management, and emotions and social structures. Rolf von Lüde is currently Head of Department of Social Sciences and Vice Dean of the School of Business, Economics and Social Sciences, University of Hamburg.