MARSSI: Model of Appraisal, Regulation, and Social Signal Interpretation

PATRICK GEBHARD, German Research Centre for Artificial Intelligence
TANJA SCHNEEBERGER, German Research Centre for Artificial Intelligence
TOBIAS BAUR, University Augsburg
ELISABETH ANDRÉ, University Augsburg

Understanding emotions of others is related to a theory of mind approach. It requires knowledge of internal appraisal and regulation processes of emotions. Multi-modal social signal classification is insufficient for understanding emotional expressions. Mainly, because many communicative emotional expressions are not directly related to internal emotional states. Moreover, the recognition of the emotional expression's direction is not considered so far. Even if social signals reveal emotional aspects, the recognition with signal classifiers cannot explain internal appraisal or regulation processes. The information the latter two provide is one approach for building cognitive empathic agents with the ability to address observations and motives in an empathic dialogue. In this paper, we introduce an emotional computational model for empathic agents. It combines a simulation of appraisal and regulation processes with a social signal interpretation that takes directions of expressions into account. Our evaluation shows that sequences of social signals can be related to emotion regulation processes. This together with appraisal and regulation knowledge enables our agent to react empathically.

Additional Key Words and Phrases: Modelling of User Emotions, Nonverbal Behavior Understanding, Empathic Agents

ACM Reference Format:
Patrick Gebhard, Tanja Schneeberger, Tobias Baur, and Elisabeth André. 2018. MARSSI: Model of Appraisal, Regulation, and Social Signal Interpretation. 1, 1, Article 4 (June 2018), 18 pages.

1 MOTIVATION

Our world is a social place. Relations with others and interaction with others are essential. In many situations, we try to understand each other yet carefully managing our mental balance. Thereby, emotions seem to play a central role [16]. Interactive agents, such as anthropomorphic robots or virtual characters, are used for training, coaching, and assistance to help people to understand each other and develop various skills [1, 18, 31, 38, 66]. The more agents are employed for social tasks; the more significant is the need for understanding user emotions, motivations, and related social behavior. All this can be exploited by interactive agents to adapt empathically to the user and the user's situation in general.

The crux of understanding emotions is that most, if not all, emotions are regulated internally [25, 65]. This is especially the case for emotions, such as shame, that are related to the appraisal of oneself [37, 60]. Only a few of the current approaches of emotion models for empathic agents take emotion regulation into account. Some of them are able to model re-appraisal processes [19, 40]. However, none of them explicitly combines a social signal interpretation with a cognitive modeling of appraisal and regulation processes.

https://doi.org/10.1145/nnnnnnn.nnnnnnn

Moreover, none of the existing recognition approaches considers the direction of emotional expressions. This means it is unclear to whom or what that emotional information applies. It is known from research in the area of emotional mimicry that the direction of emotional expressions is a crucial information to understand another's intention [12, 26]. In dyadic interactions, emotional expressions can be directed to the interaction partner, the situation, the dialog topic or at the person(s) mentioned in the utterance. By linking the gaze or head movement while observing an emotional expression, its direction can be tracked [5, 9, 61]. For example, a speaker's anger expression, directed away from the listener, provides the information that the anger is most likely addressed to something or somebody else. The knowledge about an expression's direction can be used for an automatic deduction of possible elicitors (causes) by employing different knowledge and context models. In general, the recognition of the expression's direction might be as important as the emotional expression itself, especially, if empathic agents have to generate (re-)actions based on this information.

MARSSI combines an extended social signal interpretation with a simulation of both, the appraisal and the regulation processes. The overall aim of this work is to lay the basis for a deeper analysis of social, emotional signals and their connection to cognitive processes. This may foster the widespread use of empathic agents for various assistive tasks in everyday human environments. We show a first example exploitation of our model in a job interview debriefing session. For the debriefing, a virtual character in the role of a coach addressed the observed non-verbal behavior and inferred possible appraisal and regulation hypothesis in an empathic manner.

2 RELATED WORK

2.1 Empathic Agents

Interactive systems are more likely to be accepted if the machine is aware of the user as a social actor [55, p. 247]. Furthermore, understanding how emotions work is key to social training applications [28]. In order to achieve this goal, recent developments in the area of empathic agents have initiated a shift from simple task-based human-machine interaction to a more human-like social interaction. Several approaches are addressing these requirements. Lester et al. [35] and Mulken et al. [67] are using virtual characters that are sensitive to the learners' emotional state to enhance their engagement and motivation. This is described as the persona effect. Bickmore [10, p. 131 ff.] describes the interactive fitness agent Laura that was designed to build up a relationship with a human user. In order to build a working alliance, Laura uses relational strategies like giving warm facial expression. Other approaches go further and employ cognitive models of appraisal within their systems following Wilks' argument that Digital Companions must have an understanding of the human partners' emotions as a basis for a Human-Companion relationship [74, p. 4].

Conati and Maclaren [15] present an interactive agent system that is able to model user emotions in a specific computer game. The system simulates possible user appraisals, goals, as well as motivations and models interdependencies with Bayesian networks. The emotion model uses the user's game actions as input. Rodrigues et al. [58] propose a generic computational model of empathy. In their model, they implement a reactive perception of others' affective state and the subsequent generation of an empathic response. However, Rodrigues et al. focus on the empathy between virtual agents and not between an agent and a user. Dias et al. [19] present FAtiMA, a generic and flexible architecture for emotional agents. It supports re-appraisal processes and the use of theory of mind models. How re-appraisal processes are interfering with internal situational representation is not explained.

One of the most powerful computational models of emotions is EMA. It is used by empathic agents in various systems, e.g., [63] to model appraisal and reappraisal of users [39]. Like in the previously mentioned work, goals and motivations are represented. In addition to that, EMA provides an explicit representation of coping strategies that can also be used to model a user's situational coping. Albeit coping mechanisms are related to the emotion regulation process, they differ conceptually. As a result, EMA does not allow explicit modeling of complex social emotions like shame. Also, it is unclear how to relate observed social signals to re-appraisal processes.

Looking at state-of-the-art computational models of user emotions for agents, it becomes clear that essential concepts like emotion regulation, emotional expressions direction, as well as relations to sequences of social signals, are neglected.

2.2 Emotion Modelling and Theory of Mind

Computer scientists focus on cognitive appraisal theories for emotions [44]. Because of their concept of modeling processes and signals they can be realized in computer programs. The computational modeling of emotions started in the 1980s [54] and is continuously refined [41, 59]. Psychological theories of appraisal rely on a particular input, such as, e.g., goal information, certainty, situational control, and the elicitor (who or what is the cause). Additionally, the appraisal might rely on information from a theory of mind (ToM) of others that represents hypotheses about another's mental states, status, and role [36, 56]. The outcome of the appraisal process is situational information, labeled with emotion term(s). According to the mentioned theories, elicited emotions influence behavior described with action tendencies [22], scripts [65], or facial or vocal expressions [62]. Alternatively, more general, emotions are linked to behavioral patterns how to cope with the situation [34].

Computational models realizing such theories are used to create believable behavior of virtual characters [68]. Besides, they can be used to model user's appraisal(s) in a situation. A verification of the modeled appraisal information (e.g., unexpectedness) can be realized with signal-based emotion recognition (e.g., raised eyebrow), as suggested by the psychologists Mortillaro et al. [45]. However, none of the current computational models of emotion provides this.

Currently, automatic model-based emotion recognition focuses emotional expressions and related features in voice, face, gestures, and body movements (Sec. 2.3). The essential information to whom or to what the emotion is directed, the *emotion target*, is not included in current recognition processes. Knowing, for example, that a communicated negative emotion (e.g., anger or disgust) is not directed to an interaction partner might be a relief for that partner. The results of a study by Merten [42] suggest that the aversion of gaze (by the sender) while communicating a negative emotion lets the interaction partner know this information is not directed to her/him. Also, current approaches do not consider the function of communicative emotions "[...] in dyadic interactions, as there are the speech-illustrating function [cf. [7]], the function of emotional expression, and relationship-regulation" [43]. Our model-based approach of recognizing emotional expressions takes the user's gaze and head movements into account in order to derive the emotion's target and to relate possible elicitors. Moreover, we show that the target information is central to the analysis of social signals related to emotion regulation processes.

There are few ideas in the computational realization of emotion regulation processes, mainly based on the motivation that they are an existential part of a human's emotion management. Some of the current ToM-based computational models of emotions can represent basic regulation rules (as re-appraisal rules) but not complex social emotions, such as embarrassment [39]. Also, none of the existing computational models of emotions include a real-time social signal-based emotion regulation recognition.

Recently, there are interdisciplinary approaches for computational models of emotions aiming to bridge the gap between modeled emotions and actual user emotions. One of the latest attempts employs a ToM of user emotional states in a social job interview simulation [8, 77]. Using belief, desire, and intension (BDI) rules [57], three categories of user mental states are modeled: intentions, beliefs, and emotions. The quality of social relations is based on liking and dominance values. The input of the model is the illocutionary part of speech acts (speaker intention). The model is embedded in a job interview simulation and helps to improve the system's training efficiency. A corroboration of modeled appraisal information with a real-time social signal analysis is not included.

To conclude, most of the current computational models of emotions follow the concept of cognitive appraisal-based emotion elicitation. With all existing approaches, the primary challenge remains: building a probabilistic model that relates observed social signals to possible situational appraisal regulation representations.

2.3 Social Signal Interpretation

Social signal analysis is known to be a very hard problem and a real bottleneck in social human-agent interaction. Traditionally, research has concentrated on posteriori analyses of prototypical social cues under laboratory-like conditions. Such an approach leads, however, to over-optimistic assessments of recognition rates that cannot be re-produced in naturalistic settings. A typical example includes voice data from actors for which developers of emotion recognition systems reported surprisingly high accuracy rates of nearly 80% for a seven-class problem. When moving to more naturalistic scenarios, such as child-robot interaction, accuracy rates went down considerably to about 40% for a five-class problem. An experiment that compared relevant features and recognition rates for acted and spontaneous emotions has been conducted. The experiment revealed that adequate segment lengths and relevant features could not be transferred from acted to spontaneous emotions [69].

An obvious approach to improve the robustness of the analysis is the integration of data from multiple channels. A meta-study on 30 published studies of multimodal affect detection by D'Mello and Kory comes to the interesting conclusion that performance improvement, i.e., the improvement of the fused decisions compared to the best unimodal classification, correlates significantly with the naturalness of the underlying corpus [20]. While an overall mean multimodal effect of 8.12% is reported, they also found that improvements are three times lower when classifiers are trained on natural or semi-natural data (4.39%) compared to acted data (12.1%). At first glance, the meta-study suggests that under realistic conditions there is less room for improvements than in the case of acted material. However, when analyzing the investigated approaches in more detail, it becomes apparent that most of these approaches make unrealistic assumptions, which are hard to meet in real-life environments. Therefore, they do not achieve the expected improvements different channels are combined with fixed time segments, e.g., between the beginning and the end of an utterance. It has the drawback that cues from other modalities outside the segment will be missed. Promising approaches to overcome these limitations include the use of Multi-stream Fused Hidden Markov Models [78] as well as Multidimensional Dynamic Time Warping [75].

Furthermore, attempts have been made to improve recognition rates by taking into account the dynamics of social signals. A person showing signs of happiness (usually) will not fall into a deep depression within the next few seconds. Taking the temporal context into account allows building models that are less prone to false detections. Fusion architectures based on Hidden Markov Models and Dynamic Bayesian Networks appear to be very suitable to model how social signals evolve over time. More sophisticated approaches, such as bidirectional Long Short-Term Memory [76], add more flexibility to the fusion process by learning the optimum amount of context to be taken into account.

The fusion processes mentioned above consider the temporal history of social signals. However, they do not consider the context of the social signals. So far, emotions are analyzed in isolation without considering the emotion-eliciting stimuli. This is extremely hard if not impossible [29, 53]. For example, a smile is not always a sign of happiness. People also tend to smile when feeling embarrassment [32]. Furthermore, how emotions are perceived depends on the social relationship between interlocutors [27], e.g., a person may interpret a smile of a competitor rather as gloating. Many recognition systems are not able to take these subtle differences into account. Rather they would map a smile onto the emotional state happiness. First attempts to the situational context for emotions are made by using a probabilistic framework [15]. However, this work focuses on the prediction of emotions from the situated context while the potential of external signs of emotions has not been fully exploited.

A recent study conducted by de Melo et al. analyzed the behavior of people engaged in the prisoner's dilemma with counterparts and found out that people derive information from appraisal processes when analyzing

the emotional displays of others [17]. Their study reveals the importance of appraisal-based models for the interpretation of social and emotional cues. This insight is shared by Mortillaro et al. [45]. Based on the observation that current emotion recognition systems use a so-called 'black-box' approach that map low-level features onto abstract emotion labels following statistical methods, they advocate the use of appraisal-based models to guide emotion recognition tasks. In particular, they propose appraisals as an intermediate layer between social cues and emotion labels. Nevertheless, neither the direction of emotional expressions are included, nor does the model include an estimation of emotion regulation strategies based on social cues.

3 REQUIRED CONCEPTS

Clark and Krych point out that the observation of human social signals is mandatory for a mutual understanding of a dialog partner [13]. In line with this view is the computational model of emotional grounding [11] that helps to identify the user's intention in a natural language dialogue by relying on the users's emotional state. In comparison to Conati's earlier work [14], we consider not only the emotional signals by the users but also the cause of emotions. However, both approaches did not clearly distinguish the emotion origin, such as an internal, related to a person's self, emotion (*structural emotion*), a result of the appraisal of a situation (*situational emotion*), or an emotional message expressed non-verbally (*communicative emotion*) [46, p. 111-112]. This classification schema has not found its way into computational models of emotions and approaches for recognizing emotions yet.

The combination of a social signal interpretation with modeling of structural, communicative, and situational emotions can be used to build a differentiated, probabilistic model of user's emotional states during dialogue. This approach requires a representation of (mostly) unconscious relevant processes and mental states that build a foundation for an empathic dialogue with users.

A unique, rarely by a computational model of affect included, aspect concerning structural emotions is the - mostly unconscious - regulation of *intrapersonal* emotions [64] [25, p. 6]. In that process, cultural and individual emotion regulation rules might inhibit or alter elicited structural emotions. A cognitive emotion appraisal concept extended by regulation rules enables a simulation of various adapted, or inhibited emotions. Notably, the regulation process can be related to social signals [4, 9, 47, 61], which can be recognized by a real-time social signal interpretation component. No current computational approach of emotion recognition take regulation processes and related social signals into account. Both, their importance and necessity for understanding human emotions are described by the cognitive psychoanalysts Moser and von Zeppelin [48, 49]. Relying on the combination of regulation processes and social signals for emotion recognition is of particular importance when considering that the mapping of emotional expression (even considering the fusion of several modalities) onto emotional states is not reliable [26, 29, 30, 53, 71].

3.1 Structural, Situational, and Communicative Emotions

In 1990 the psychologists Bänninger-Huber et al. introduced a ToM concept of how to combine an offline social signal interpretation with modeled emotions and emotion regulation processes [4]. The work aimed at the creation of an emotion regulation process model. Based on this, Moser and von Zeppelin designed a theory of emotions that differentiates between *communicative emotions*, *structural emotions*, and *situational emotions* [47].

This *functional classification of emotions* helps to describe emotions and their implications on internal processes as well as their reflection in behavior more distinguishable:

Structural emotions represent information about the appraisal of oneself and hence are related to the self-image (Fig. 1, top, left and right). Such emotions are, e.g., shame, pride or gratitude.

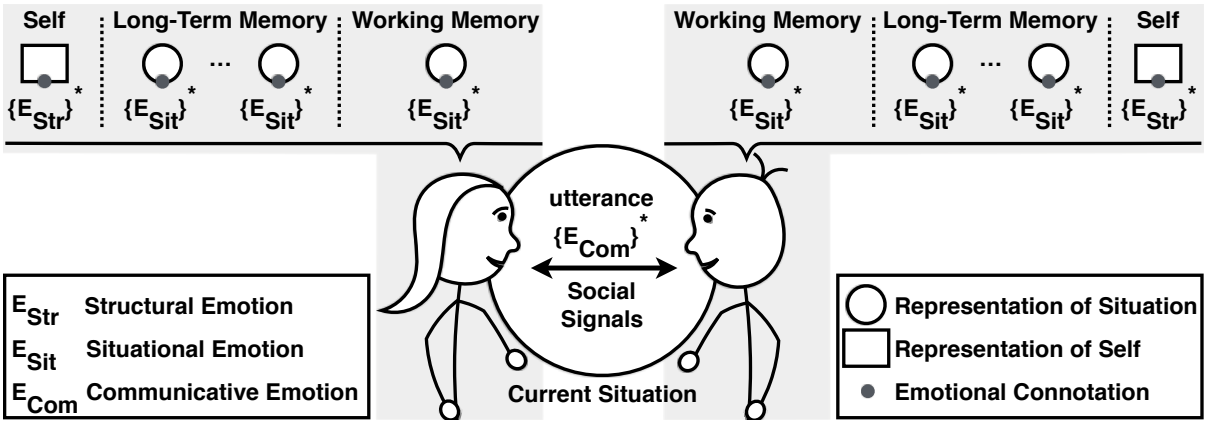


Fig. 1. Structural emotions, situational emotions, and communicative emotions in a dyadic interaction setup.

Situational emotions represent information that is linked to a topic or situation that have been experienced (Fig. 1, top, center, long-term and working memory). Situational emotions reflect the level of security. More specific, such emotions like, e.g., fear or distress reflect the fact that the situation comes with unforeseen or unbearable requirements. If a situation addresses social skills or relations, the emotions shame or pride might be linked.

Communicative emotions are encoded non-verbally in *sequences of social signals*, like in vocal or facial expressions (Fig. 1, center). They are, e.g., described by Ekman [21]. "Communicative affects bring the regulatory systems [and related structural, and situational emotions, author's remark] of both interaction partners in relation and they provide rapid information about the partner's regulatory state." [47, p. 111]. One of the most crucial aspects of communicative emotions is that they are directed towards the dialog partner or situational objects [61] [3, p. 118 ff.]. The class of communicative emotions includes social signals that are used for relationship regulation/management (esp. smile) [3, p. 72 ff.], which is related to social mimicry processes [26, 33].

3.2 Emotion Regulation

An emerging research focus on cognitive emotion theories is the *regulation of emotions* [25]. Tomkins proposed that adult emotions are almost always regulated [65]. The regulation of emotions describes the process of suppressing or changing emotions if they do not fit the current individual situation. The main purpose of the regulation process is to "cover" an unwanted emotion with others in order to (re-)establish the feeling of being secure [64].

The regulation process changes the situational appraisal information, which elicits a different emotion reflecting a "better" (with regard to the individual's situational appraisal) management (coping) of the situation. The employed *regulation strategy* changes situational values of individuals' internal situational representation in the working memory (Fig. 1, top). Classes of situational changes are described by Moser [46, p. 39]: 1) *actor transformations (self as actor → other as actor, other as actor → self as actor*), 2) *action transformations (e.g., action → opposite of action, action → denial of action)*, and 3) *object transformations (object x → self as object, object x → y as object, x≠y, y≠self).* As a result, an individual situational representation differs from the current outside situation. This view explains different individual situational descriptions. With our approach, we follow Moors et al.'s suggestion that the regulation of emotion should be part of any appraisal process model [44].

There is evidence that the regulation process can be observed through related social signals [3, 9, 47, 51, 61]. For building a computational ToM model for the structural emotion shame, we rely on Nathanson's shame

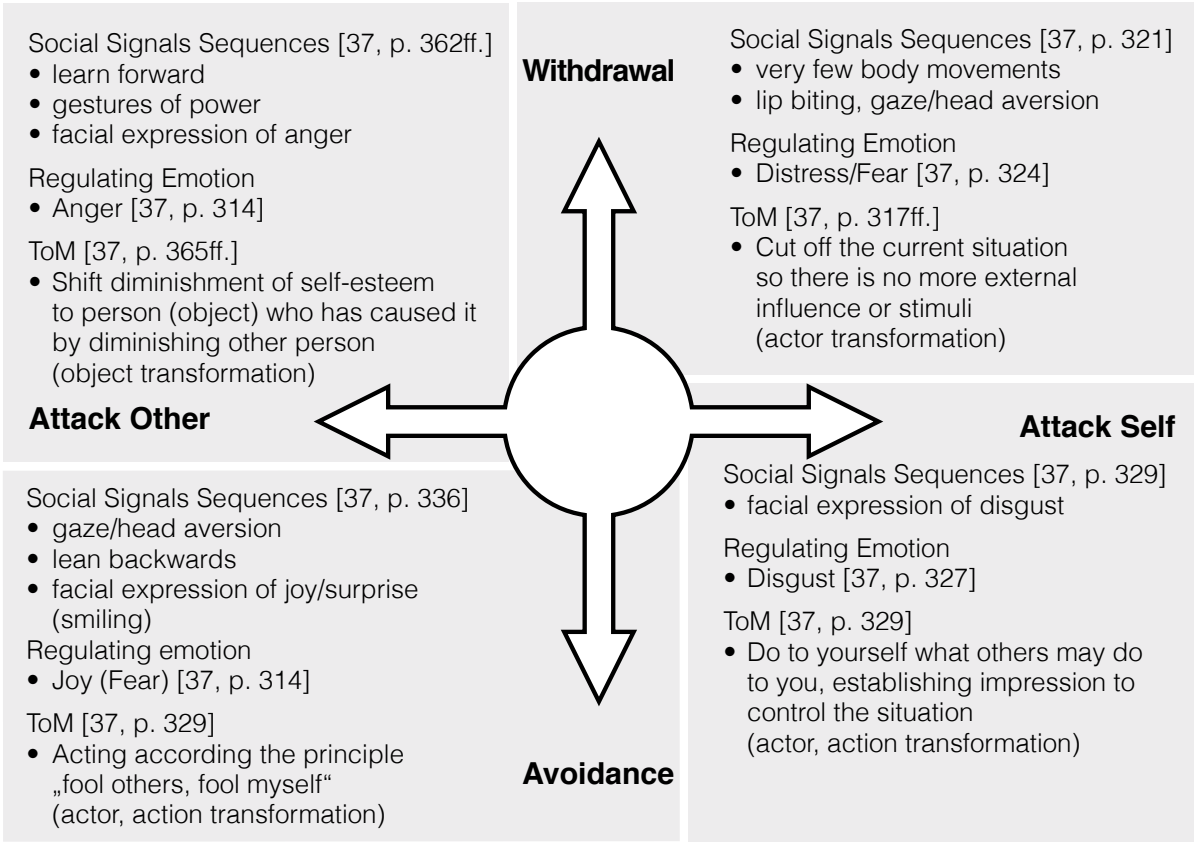


Fig. 2. Possible shame regulation strategies, related sequences of social signals, and explanation examples.

regulation model [51]. It is a model that takes 1) clinical observations, 2) individual background and information about personal motivations, and 3) typical sequences of social signals of emotion regulation into account. For the regulation of the structural emotion shame, Nathanson describes four strategies with related social signals, and regulating emotions: 1) *Avoidance*, 2) *Attack Self*, 3) *Attack Other*, and 4) *Withdrawal* (Fig. 2). Most likely, regulating emotions are expressed (as a communicative emotion) in the sequence of social signals that is related to individually chosen regulation strategy.

For example, *Withdrawal* is accompanied by head adaptors, lip biting, slight body movements, or avert head/gaze, *Avoidance* is accompanied by averting head/gaze or gaze wandering. Social signals are indicating a regulation process sometimes differ only minimally. For *Attack Other*, related social signals are directed gaze, spacious gestures/posture. Both, 1) the social signals of the regulation process (while processing the regulation strategy), and 2) the social signals of the regulating emotion compose identifiable signal patterns. These patterns allow conclusions to be drawn on the regulatory process and strategy. In the case of *Avoidance*, the regulating emotion is joy (triggered by the concept *"fool others fool myself"*, [51, p. 339]) with the corresponding facial expression smile. These signal sequences can be detected and interpreted in real-time by the MARSSI's social signal interpretation component. A result is an increased accuracy for recognizing structural emotions.

4 MARSSI

This section discusses required knowledge representation, the components, and the overall workflow of MARSSI. The simulation of possible user emotions relies on cognitive modeling of appraisal rules, emotion regulation rules, and social signal classifiers. The latter requires real-time signal data from an eye tracker for capturing eye movement, a depth camera for capturing head movement, facial expression, gestures, and posture; and a microphone for voice.

4.1 Emotion Classes, Rules, and Classifiers

MARSSI extends the emotion types from Orthony, Clore, and Collins (OCC) [52, p. 19 ff.] by Moser's and von Zeppelin's functional emotion classification (Sec. 3.1). All OCC emotions are assigned to the functional emotion class situational emotion, except the emotions of the types *Attribution* and *Well-Being/Attribution.* They are assigned to the functional class structural emotions since they are related to the self-image.

An *Appraisal Rule* defines how a situation is judged. With regard to cognitive appraisal theories, the situation is the elicitor of emotion. An appraisal rule represents how a user would appraise a situation. Multiple appraisals are allowed. We rely on the OCC appraisal theory [52] with its implementation by A Layered Model of Affect (ALMA) [23, 24], e.g. *GoodActSelf* → {*agency=self, praiseworthiness=1.0*}. In this work, we use ALMA'S appraisal tag representation, e.g., GoodActSelf, to describe an appraisal. In this case, the tag is a shortcut to the reasoning process in which appraisal rules infer a positive praiseworthiness of the action regarding the agent's goals, current situation, and related facts. MARSSI extends the appraisal notation with a confidence value representing a value how likely the appraisal fits the detected social signals. The value is computed by social signal classifiers.

A *Regulation Rule* defines how an internal emotion is regulated by changing the current appraisal information triggering a re-appraisal process that elicits a regulating emotion. Regulation rules are used to model how a user might regulate internal emotions. Multiple regulations are allowed. MARSSI extends ALMA by processing regulation rules (Sec. 3.2). We created regulation rules for the structural emotion shame following Nathanson's regulation theory (Fig. 2). All regulation rules contain *situational change rules* (marked with *sit_chg*) and corresponding OCC appraisal information: 1) *AttackOther* → {*sit_chg:object self* → *object other; agency = other, praiseworthiness = -1.0*}. This rule regulates shame with reproach, elicited by a negative praiseworthiness by shifting the appraisal focus from one own's flaw to a blameworthy action of the person who is responsible for the shame experience. 2) *Withdrawal* → {*sit_chg:other as actor* → *self as actor; agency = self, desirability = -1.0*}. This rule regulates shame with distress, elicited by a negative desirability but replacing the person who is responsible for the shame experience with oneself, to the purpose of having control over the situation. A similar Withdrawal rule might include a negative likelihood to elicit the regulating emotion fear. 3) *Avoidance* → {*sit_chg:action* → *opposite of action*|*denial of action*|*...; agency = self, desirability = 1.0*}. This rule regulates shame with joy, elicited by a positive desirability of the imagined positive event in which the shame action has not happened. 4) *AttackSelf* → {*sit_chg:other as actor* → *self as actor, action* → *intellectualization of action; agency = self, liking = -1.0*}. This rule regulates shame with disgust, elicited by a negative liking and the transformation of the shameful action into an own "ugly" character feature that is less intense and can be changed by oneself in the future. Because the person who is responsible for the shame experience is replaced with oneself implicates having control over the situation. All regulating emotions of the shame regulation rules are situational emotions that are most likely communicated (non-)verbally (e.g., [51, p. 315 ff.]), hence become communicative emotions. Note that each regulation rule's OCC variable hold the maximal value (e.g., 1.0 or -1.0). Its sign determines the type of emotion. Its value can be used to calculate an emotion's intensity. Currently, we are interested in the type only. Each rule holds a confidence value that is computed by social signal classifiers during runtime, representing a value how likely the regulation fits the detected social signals.

Social Signal Classifiers in MARSSI are conceptually related to appraisal and regulation information expressed as communicative emotions. We employ classifiers that are able to detect *sequences of social signals* as they occur in the situation of emotion regulation. We focus on classifiers for head (gaze), specific gestures, and posture changes for the following appraisal and regulation information: 1) *BadEvent*: user expresses anger directed towards the situation - away from the dialog partner, 2) *BadActOther*: user expresses anger towards the dialog partner, 3) *BadActSelf*: user shows facial expression of shame (e.g., blushing), head/gaze points downwards, posture is slumped down, for all shame regulation classifiers: the regulation takes time and might be accompanied by 4) *BadActSelf* → *AttackOther*: a lean forward posture and gestures that take up room, and the user expresses anger towards the dialog partner, 5) *BadActSelf* → *Avoidance*: a lean back posture, gaze and head aversion, and the user expresses joy towards the dialog partner, 6) *BadActSelf* → *Withdrawal*: few body movements, gaze/aversion, and the user expresses fear away from the dialog partner, 7) *BadActSelf* → *AttackSelf*: expresses disgust away from the dialog partner, head/gaze is mainly pointed downward.

To this end, the models for recognizing single social cues included in MARSSI are trained using machine-learning supported annotation tool NOVA[1]. To fuse multiple social signals, we employ *Dynamic Bayesian Networks* (*DBNs*) [50]. One of their main advantages is that they allow theory-based modeling of the structure and relevant features (represented by nodes) of a higher-level concept (e.g., regulation of shame with Withdrawal), but the probability distribution of single nodes may be learned from data. Further DBNs support the concept of time, allowing to model and learn temporal sequences for the interpretation of social signals. We first employ multiple classifiers trained to predict single social cues (such as facial expressions, gaze direction) to create automated annotations. For each situation, human experts manually label higher-level concepts, such as the emotion regulation strategies (Sec. 5).

During run-time, a confidence value, computed by the output of the nonverbal interpretation of the appraisal and regulation strategy is forwarded to the emotion simulation component, updating the possibilities of each modeled appraisal and regulation information.

4.2 Components and Workflow

Figure 3 shows how MARSSI (bottom) extends a typical appraisal approach (top) illustrating the components and workflow. Both approaches are extended by a Social Signal Interpretation component.
The MARSSI user emotion simulation is based on ALMA [23] and the Social Signal Interpretation framework (SSI) [73]. ALMA provides a flexible appraisal interface and is able to simulate multiple emotional states in parallel. It was extended straightforwardly by the required regulation process and required confidence representations for appraisal and regulation representation. SSI especially allows the synchronized processing of multiple sensor inputs in real-time. This includes the extraction of relevant features at runtime and the appliance of machine learning models, such as deep neural networks or support vector machines (SVM) for predicting single cues, such as changes in gaze direction, facial expressions, gestures, and postures.

Our simulation of user emotions is structured according to conceptually coherent situations in dyadic interactions (e.g., question-answer, or comment) between a speaker and a listener. Technically, we rely on a voice signal analysis (plus gaze and head movement detection) to infer the dialog partner's attention, and actions (e.g., a user starts/stops speaking) implemented as SSI classifiers [6]. The speaker is supposed to ask an emotion triggering question. While the speaker starts asking the question, the simulation of the listener's emotions is prepared (preparation phase), and the signal recognition is activated (recognition phase).

The preparation phase triggers the actual emotion simulation by a *set of appraisal and regulation annotation* given as input (e.g., {*([BadActSelf], [AttackOther, Avoidance, Withdrawal, AttackSelf])*}). Currently, the annotation is provided by human experts that annotate the situation with that specific information (Sec. 5). The annotation

[1] http://github.com/hcmlab/nova

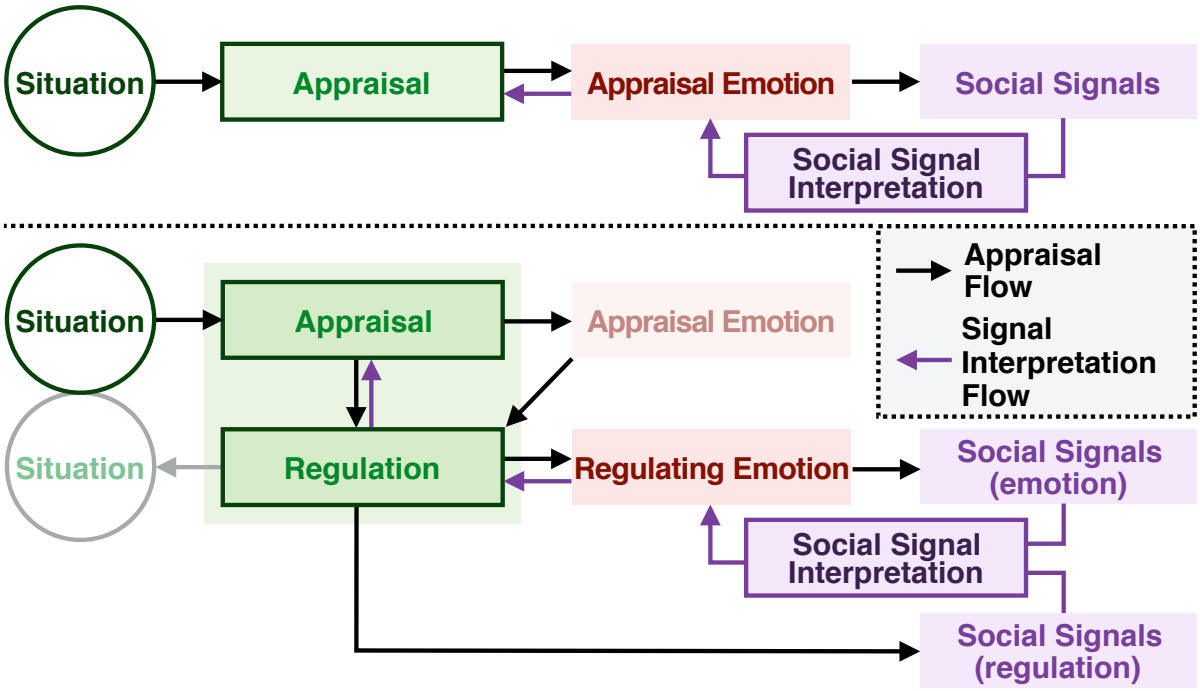


Fig. 3. Typical cognitive appraisal process flow (top), MARSSI appraisal and regulation flow (bottom).

could, theoretically, derived automatically having a full-blown ToM of that specific user. In this work, we focus on the simulation of the interconnections between appraisal, regulation, and social signals (Sec. 3.1). Each appraisal and regulation rule input let MARSSI create a separate *emotion simulation session* (*emo_ss*). The example input creates five emo_ss, each holding appraisal information, the *elicited emotion*, and (if a regulation rule is stated) the *regulation rule*, and the *regulating emotion*: 1) (*BadActSelf* → *Shame*), 2) (*BadActSelf* → *Shame* → *AttackOther* → *Reproach*), 3) (*BadActSelf* → *Shame* → *Avoidance* → *Distress*), 4) (*BadActSelf* → *Shame* → *Withdrawal* → *Joy*), 5) (*BadActSelf* → *Shame* → *AttackSelf* → *Disgust*).

The recognition phase lasts as long as the listener handles the question or the comment. Within that phase, the Social Signal Interpretation updates the appraisal and regulation confidence values in each emo_ss reflecting the match of detected social signals to the appraisal and regulation information in each emo_ss.

5 EVALUATION AND EXAMPLE SIMULATION

This section explains how we employed MARSSI for an empathic agent. First, we need recorded data of participants in specific situations that elicit the structural emotion shame to build our corpus. We used a job interview situation and tried to elicit the structural emotion shame in the interviewees. To generate shame eliciting situations, we conducted a pre-study. Two job coaching experts identified six possible shame eliciting situations considering Nathanson's work (Sec. 3.2). 26 participants (age 18 - 29, M = 21.71, SD = 2.91) were asked to put themselves into a position of a job applicant experiencing these six different situations. The task of the participants was to describe in their own words how they would react. The answers were analyzed by two psychologists and assigned to Nathanson's four shame regulation strategies (Fig. 2). Finally, we identified five situations that elicit

the structural emotion shame, e.g., *"Before we begin, let me ask a short question: Where did you find your outfit? It really doesn't suit you."*

To generate our corpus, we created a 15min job interview with the five shame eliciting situations from the pre-study. In our evaluation, this job interview was conducted by a female interviewer with 20 participants (10 female, age 19 - 30, M = 24.60, SD = 4.08) as a role-play. After welcoming the participants, they were asked to imagine that they applied for a student assistant job in their favorite faculty. Each participant is sent to the interviewer's office for a job interview. Afterwards, the participant answered demographic questions and was compensated. The interviews were recorded with a depth camera and a head-mounted microphone.

In total, 100 (20 participants in five situations) shame eliciting situations are building the corpus for the analysis. We annotated the obtained data in order to create the social signal classifiers. Each situation was classified independently by three students, that were not related to the experiment neither knew about the aim of the study. They were trained beforehand to classify Nathanson's four shame regulation strategies. Overall, 300 labels were assigned as follows: 83 Withdrawal, 105 Attack Self, 98 Avoidance and 14 Attack Other. For assessing the reliability of agreement Fleiss' kappa was calculated for three raters, four labels, and 100 data points. With 0.7301 it is considered as substantial agreement.

Based on this data, we trained the Bayesian network in a 50:50 split validation approach. To this end, we employed several social signal processing algorithms to generate labels for single social cues on multiple modalities of both the interviewer and the candidate. Some cues are calculated based on single, meaningful features, such as the energy of the motion vectors of both hands of a participant or the overall movement of the hands, head touches, and the openness of the body posture [6].

For more complex cues, e.g., subtle smiles, we employed an SVM to train models based on manual annotations on the training subset of our corpus. For cues related to the head and face, we thereby extracted OPENFACE [2] features. Analogously, we repeated this step for other modalities, such as the paralinguistic channel, by training a model to detect spoken words, fillers, and silence, as well as models to detect the level of arousal from the audio modality based on GEMAPS [70] features. A human annotator interactively corrected the annotations when necessary, and after each session, the models have been retrained as proposed in [72].

To find the ground truth of the observed emotion regulation strategy, we additionally labeled time segments including the duration of each question and the candidate's answer, with 1) the type of question as additional context information and 2) with the rating of human labelers for the classes related to regulation cues (e.g., AttackOther, AttackSelf, Avoidance, Withdrawal, and None).

Finally, based on these semi-automated annotations we created a training set. It contains the parallel appearance of the ground truth labels for the shame emotion regulation strategy, the context information and the single observed social cues (we discretized continuous annotations) and trained a DBN using the Expectation Maximization algorithm, to learn both the distribution of the single labels in our corpus, but also their influence on the single shame regulation strategies. Overall, the network achieved a precision of 82% for Avoidance, 65% for AttackSelf and 64% for Withdrawal from non-verbal behaviors only. The training data provides too few social signals related to the AttackOther strategy. As a result, the DBN could not be trained to that extend.

In a next step, we used the cognitive modeling and the trained social signal classifiers to simulate user emotions in real-time in a debriefing session with our interactive virtual character Tom. He has the role of a coach discussing the user's (non-verbal) reaction to the interviewer's question. Tom is embedded in a 3d virtual environment (Fig. 5) capable of performing social cue-based interaction with the user. He is able to perform lip-sync speech output using the state-of-the-art Nuance Text-To-Speech system. Tom comes with 36 conversational motion-captured gestures and has 14 facial expressions including the six basic emotion expressions.

For each shame question, possible appraisals and regulations of the applicant were prepared by MARSSI. Each preparation phase (Sec. 4.2) is triggered by the voice activity signal of the job interviewer, posing the question. In

fact, the following appraisal/regulation input is given to MARSSI for each shame question: {*([BadEvent]), [BadActOther], ([BadActSelf], [AttackOther, Avoidance, Withdrawal, AttackSelf])*} with *[BadEvent]* denotes the appraisal that the situation as noisy, *[BadActOther]* denotes the appraisal that the interviewer's action is blameworthy, e.g., the interviewer speaks with an inappropriate low voice, and *[BadActSelf]* denotes the appraisal that the question triggers a blameworthy memory of the applicant. The latter elicits the structural emotion shame that the applicant most likely will regulate with the 4 mentioned regulation strategies (Sec. 3.2 and Sec. 4.1). As a result seven emo_ss (Sec. 4.2) are created holding appraisal information, the *elicited emotion*, and (if a regulation rule is stated) the *regulation rule*, and the *regulating emotion*: 1) (*BadEvent* → *Distress*), 2) (*BadActOther* → *Reproach*), 3) (*BadActSelf* → *Shame*), 4) (*BadActSelf* → *Shame* → *AttackOther* → *Reproach*), 5) (*BadActSelf* → *Shame* → *Avoidance* → *Distress*), 6) (*BadActSelf* → *Shame* → *Withdrawal* → *Joy*), 7) (*BadActSelf* → *Shame* → *AttackSelf* → *Disgust*). At the same time, the related social signal classifiers are activated (Sec. 4.1). At runtime, the confidence values from the classifiers update appraisal and regulation representations (Fig. 4).

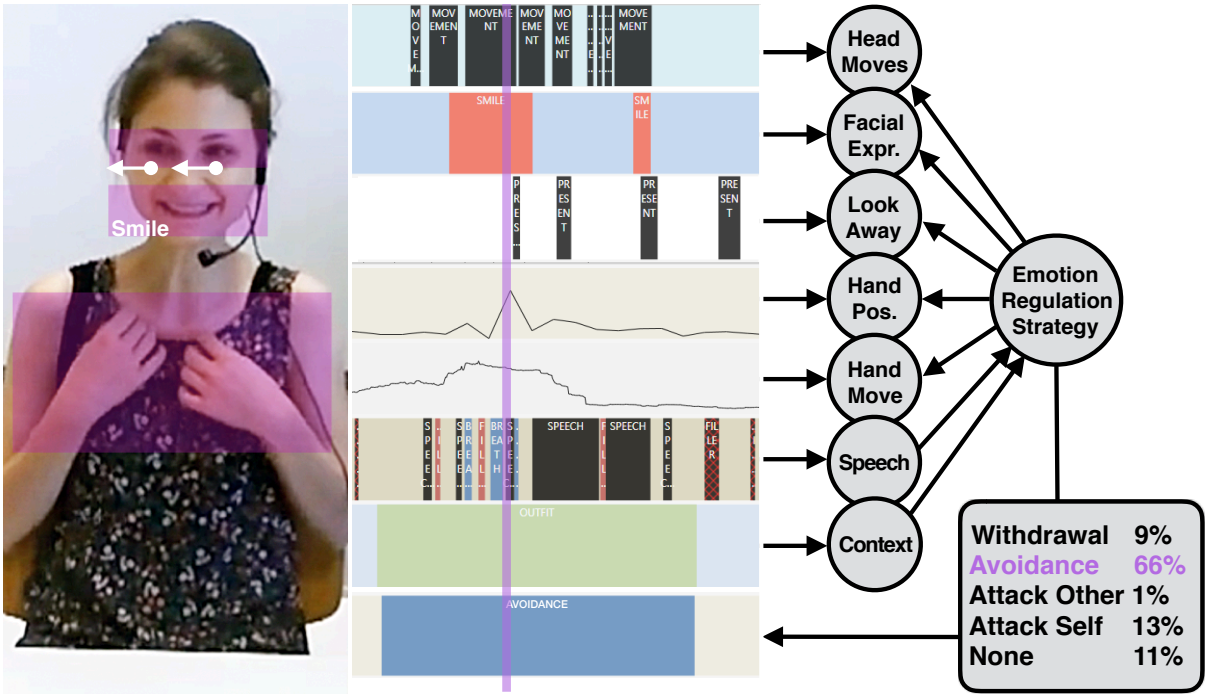


Fig. 4. Recognized and annotated cues are fed in a DBN that infers the current shame regulation strategy and predicts it in real-time.

Our empathic agent exploits MARSSI's knowledge of the appraisal and the regulation strategies in order to generate an empathic reaction. Currently, the reaction is based on the detected appraisal or regulation with the highest confidence value. The aim is to support in the user's self-reflection by explaining to her what MARSSI discovered from the social signals. We elucidate this with the example of the regulation strategy Avoidance. Avoidance is one of the four regulation strategies when experiencing the structural emotion shame [51]. It is accompanied by specific facial expressions and body language (Sec. 3.2). This strategy can also be expressed verbally by redirecting the subject to another. We focus on the facial expression and body language. In general, Tom (Fig.5, right) would first explain what social signals MARSSI have detected and which regulating emotions

are related. Afterwards, he would subtly explain the connection to the underlying structural emotion. We want to outline a possible interaction between a user and the coach where MARSSI detected the following rule *Avoidance* → {*sit_chg:action* → *opposite of action|denial of action|...; agency = self, desirability = 1.0*} in the example situation with the interviewer *"Before we begin, let me ask a short question: Where did you find your outfit? It really doesn't suit you.* This rule regulates shame with joy, elicited by a desirable imagined positive event in which the shame action has not happened.

As seen in Tom's explanation, he does not directly address the structural emotion. Especially in those cases where the underlying structural emotion might be shame, the subtle approach is extremely important. Since shame is the emotion that is connected to the evaluation of the self, the coach has to be very sensitive such that the user is still able to preserve his self [37, 60].

In the example situation, MARSSI recognized the regulation strategy Avoidance. We generate the explanations with textual templates for: 1) situation description (and for the first shame question, explanations of Tom's role) and found social signal sequences related to appraisal and regulation strategies (Fig. 5, 1), 2) general explanation how such signals could have interpreted (Fig. 5, 2), and 3) explanation of the regulation process and typical observations (Fig. 5, 3), which we took from descriptions of Nathanson [51, p. 303 ff.] and the two coaching experts.

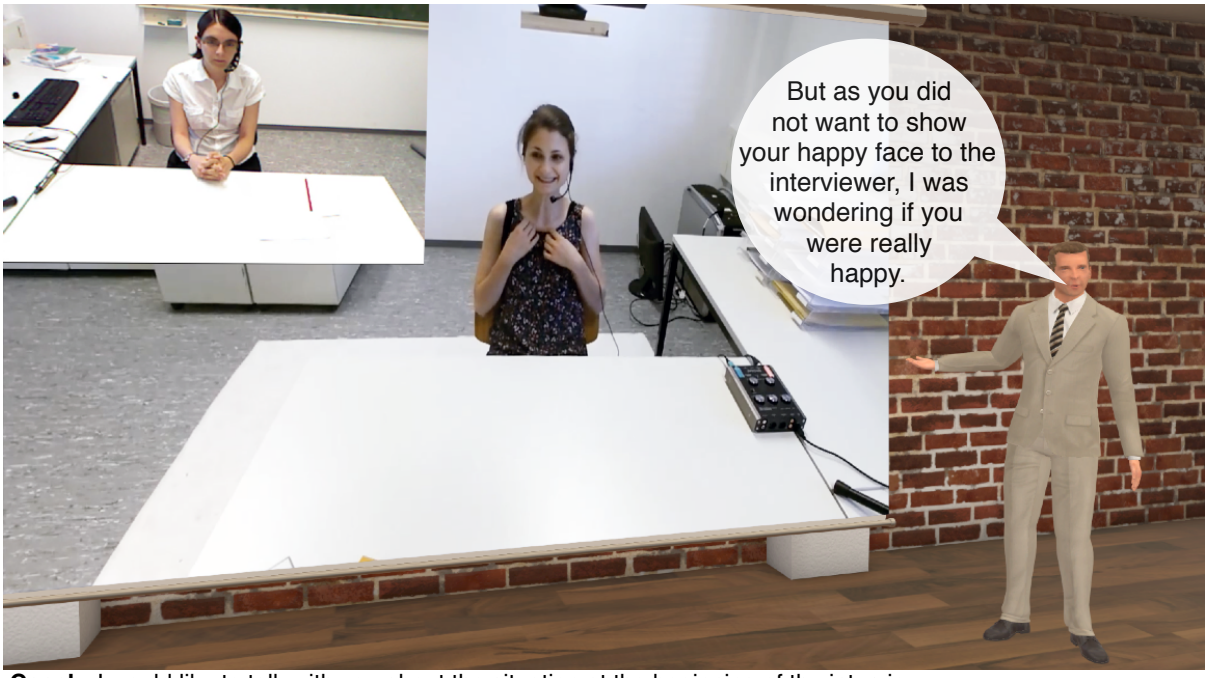
6 CONCLUSION AND FUTURE WORK

In this paper, we have presented the computational model of emotion MARSSI that relates appraisal rules and emotion regulation rules with social signal interpretation. MARSSI employs an extended theory of emotions that comes with three functional dimensions to emotions: communicative emotions, situative emotions, and structural emotions. This notation allows a more precise description of emotions. Also, it allows defining possible, plausible relations between communicative emotions (cf. emotional expressions) and sequences of social signals to individual appraisal and regulation strategies. The latter can be triggered by elicited structural emotions, such as shame, which was our focus in this work.

On a conceptual level, the implications of MARSSI are twofold: 1) advancement of social signal classifiers with regard to an improved recognition of emotional aspects that can be related to structural emotions and 2) explanation of detected communicative emotions based on represented appraisal and regulation strategies and confidence values that are derived by the advanced social signal classifiers. The advancement of social signal classifiers is achieved by learning time and spatial relations of social signal sequences that are related to internal appraisal and regulation processes for a specific context. This process especially takes head and eye movements during communicative emotions into account reflecting the so far neglected aspect that human emotional expressions are directed. The MARSSI appraisal and regulation strategies allow possible explanations of detected communicative emotions concerning internal motivations. They are represented within the strategies and derived by related theories of emotion regulation.

We used a corpus-based approach to create our social signal classifiers in the context of job interviews. Some of the job interview questions are designed to elicit the structural emotion shame. Using MARSSI, we were able to model appraisal and regulation strategies that might occur in an applicant during a job interview. In a debriefing session, we used this knowledge together with our advanced social signal classifiers for analyzing each individual's social cues and for computing confidence values for modeled regulation strategies. An empathic virtual agent in the role of a job interview coach explains the regulation strategy with the highest confidence value. This enables the virtual coach to empathically address the possible elicited structural emotion shame explaining further details about the detected social cues.

MARSSI is a starting point for various types of research. The modeling of regulation strategies can be extended to cover other structural and even situational emotions. The notation of situational emotions could be exploited



Coach: I would like to talk with you about the situation at the beginning of the interview. The interviewer commented on your outfit. Is this ok with you?

User: Sure.

Coach: Do you first want to see the video from the interviewer's position?

User: Yes.

[system plays the recorded video, pauses three times, coach explains …]

Coach: In this situation, the interviewer was attacking your outfit saying that it does not fit you.

1 As you know, I kept a watch on your facial expression and your body language during the interview. I could observe that you were smiling and looking away from the interviewer while answering.

Coach: It seems like you did not want to look at the interviewer anymore though you were smiling. Because of the smile, I could have thought you were happy first. But as you did not want to show your happy face to the interviewer, I was wondering if you were really happy. Maybe the attack on your appearance made you feel bad, but you did not want to show it. That is ok.

Coach: To defend themselves, others sometimes do not at all understand the attack but think the interviewer said their outfit fitted nicely. If someone said my suit didn't look good, I also would feel hurt. But don't worry, the interviewer just said this to get you off your feet, because you are already at the advanced level of the training.

Fig. 5. Virtual coach discusses prominent situations.

to learn how users emotionally remember a specific situation. An empathic agent might observe in the non-verbal behavior of users if past job interviews went bad. Since the advanced social signal classifiers rely on context information, we have to investigate if such classifiers can be applied in other contexts than the used job interview context. One important issue is the acceptance of such agents, especially if they can discuss their observations with the user. This could be exploited for agents to learn individual regulation patterns to refine the user model.

ACKNOWLEDGMENT

This work is partially funded by the German Ministry of Education and Research (BMBF) within the EmpaT project (funding code 16SV7229K). The authors thank the colleagues Thomas Anstadt, Ulrich Moser, and Eva Bänninger-Huber for providing their expertise in various phases of the presented work. The authors thank the Charamel GmbH and the TriCAT GmbH for realizing our requirements with regard to the virtual agents and the 3d environment.

REFERENCES

[1] Keith Anderson, Elisabeth André, Tobias Baur, Sara Bernardini, Mathieu Chollet, Evangelia Chryssafidou, Ionut Damian, Cathy Ennis, Arjan Egges, Patrick Gebhard, et al. 2013. The TARDIS Framework: Intelligent Virtual Agents for Social Coaching in Job Interviews. In *Advances in Computer Entertainment*. Springer, 476–491.

[2] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. Openface: an open source facial behavior analysis toolkit. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. IEEE, 1–10.

[3] Eva Bänninger-Huber. 1996. *Mimik Übertragung Interaktion: Die untersuchung Affektiver Prozesse in der Psychotherapie*. Huber.

[4] Eva Bänninger-Huber, Ulrich Moser, and Felix Steiner. 1990. Mikroanalytische Untersuchung affektiver Regulierungsprozene in Paar-Interaktionen. *Zeitschrift für klinische Psychologie* 19, 2 (1990), 123–143.

[5] Eva Bänninger-Huber and Felix Steiner. 1992. Identifying Microsequences: A New Methodological Approach to the Analysis of Affective Regulatory Processes. In *"Two Butterflies on My Head..."*. Springer, 257–276.

[6] Tobias Baur, Gregor Mehlmann, Ionut Damian, Florian Lingenfelser, Johannes Wagner, Birgit Lugrin, Elisabeth André, and Patrick Gebhard. 2015. Context-Aware Automated Analysis and Annotation of Social Human–Agent Interactions. *ACM Trans. Interact. Intell. Syst.* 5, 2, Article 11 (June 2015), 33 pages.

[7] Janet Beavin Bavelas and Nicole Chovil. 1997. *The psychology of facial expression*. Cambridge University Press, Cambridge, U.K., Chapter Faces in dialogue, 334–346.

[8] Marwen Belkaid and Nicolas Sabouret. 2014. A logical model of Theory of Mind for virtual agents in the context of job interview simulation. *arXiv preprint arXiv:1402.5043* (2014).

[9] Cord Benecke. 2002. *Mimischer Affektausdruck und Sprachinhalt*. Ph.D. Dissertation. Saarland University.

[10] Timothy Wallace Bickmore. 2003. *Relational Agents: Effecting Change through Human-Computer Relationships*. Ph.D. Dissertation. Massachusetts Institute of Technology.

[11] Wauter Bosma and Elisabeth André. 2004. Exploiting Emotions to Disambiguate Dialogue Acts. In *Proceedings of the 2004 International Conference on Intelligent User Interfaces, January 13-16, 2004, Funchal, Madeira, Portugal*. 85–92.

[12] Patrick Bourgeois and Ursula Hess. 2008. The impact of social context on mimicry. *Biological Psychology* 77, 3 (2008), 343 – 352. https://doi.org/10.1016/j.biopsycho.2007.11.008

[13] Herbert H. Clark and Meredyth A. Krych. 2004. Speaking while monitoring addressees for understanding. *Journal of memory and language* 50, 1 (2004), 62–81.

[14] Cristina Conati. 2002. Probabilistic Assessment of User's Emotions in Educational Games. *Applied Artificial Intelligence* 16, 7-8 (2002), 555–575.

[15] Cristina Conati and Heather Maclaren. 2009. Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction* 19, 3 (2009), 267–303.

[16] Antonio Damasio and Sebastian Vogel (Translator). 2017. *Am Anfang war das Gefühl (original title: The Strange Order of Things: Life, Feeling, and the Making of Cultures)*. Siedler, München.

[17] Celso M. de Melo, Peter J. Carnevale, Stephen J. Read, and Jonathan Gratch. 2014. Reading people's minds from emotion expressions in interdependent decision making. *Journal of personality and social psychology* 106, 1 (2014), 73.

[18] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, and Louis-Philippe Morency. 2014. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems (AAMAS '14)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1061–1068. http://dl.acm.org/citation.cfm?id=2617388.2617415

[19] Joao Dias, Samuel Mascarenhas, and Ana Paiva. 2014. FAtiMA Modular: Towards an Agent Architecture with a Generic Appraisal Framework. In *Emotion Modeling*. Springer, 44–56.

[20] Sidney K. D'Mello and Jacqueline M. Kory. 2012. Consistent but Modest: A Meta-Analysis on Unimodal and Multimodal Affect Detection Accuracies from 30 Studies. In *ICMI*, Louis-Philippe Morency, Dan Bohus, Hamid K. Aghajan, Justine Cassell, Anton Nijholt, and Julien Epps (Eds.). ACM, 31–38.

[21] Paul Ekman. 1992. An Argument for Basic Emotions. *Cognition & Emotion* 6, 3-4 (1992), 169–200.

[22] Nico H. Frijda. 1987. *The Emotions (Studies in Emotion & Social Interaction).* Cambridge University Press.

[23] Patrick Gebhard. 2005. ALMA - A Layered Model of Affect. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, Frank Dignum, Virginia Dignum, Sven Koenig, Sarit Kraus, Munindar P. Singh, and Michael Wooldridge (Eds.). ACM, 29–36.

[24] Patrick Gebhard, Michael Kipp, Martin Klesen, and Thomas Rist. 2003. Adding the emotional dimension to scripting character dialogues. In *Proceedings of the 4th International Workshop on Intelligent Virtual Agents (LNAI 2792)*, Jaime G. Carbonell and Jörg Siekmann (Eds.). Springer, Berlin Heidelberg, 48–56.

[25] James J. Gross. 2013. *Handbook of Emotion Regulation.* Guilford publications.

[26] Ursula Hess and Agneta Fischer. 2013. Emotional mimicry as social regulation. *Personality and Social Psychology Review* 17, 2 (2013), 142–157.

[27] Ursula Hess and Shlomo Hareli. 2015. The influence of context on emotion recognition in humans. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. 03. 1–6. https://doi.org/10.1109/FG.2015.7284842

[28] W. Lewis Johnson, Jeff W. Rickel, and James C. Lester. 2000. Animated Pedagogical Agents: Face-to-face Interaction in Interactive Learning Environments. *International Journal of Artificial intelligence in education* 11, 1 (2000), 47–78.

[29] Markus Kächele, Stefanie Rukavina, Günther Palm, Friedhelm Schwenker, and Martin Schels. 2015. Paradigms for the Construction and Annotation of Emotional Corpora for Real-World Human-Computer-Interaction. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM)*. SciTePress, 367–373.

[30] Susanne Kaiser and Thomas Wehrle. 2001. The Role of Facial Expression in Intra-individual and Inter-individual Emotion Regulation. *Emotional and intelligent II: The tangled knot of cognition* (2001), 61–66.

[31] Ashish Kapoor and Rosalind W. Picard. 2005. Multimodal Affect Recognition in Learning Environments. In *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*. ACM, New York, NY, USA, 677–682. https://doi.org/10.1145/1101149.1101300

[32] Dacher Keltner. 1995. Signs of Appeasement: Evidence for the Distinct Displays of Embarrassment, Amusement, and Shame. *Journal of Personality and Social Psychology* 68, 3 (1995), 441.

[33] Jessica L. Lakin, Valerie E. Jefferis, Clara Michelle Cheng, and Tanya L. Chartrand. 2003. The Chameleon Effect as Social Glue: Evidence for the Evolutionary Significance of Nonconscious Mimicry. *Journal of nonverbal behavior* 27, 3 (2003), 145–162.

[34] Richard S. Lazarus. 1991. *Emotion and Adaptation.* Oxford University Press.

[35] James C. Lester, Sharolyn A. Converse, Susan E. Kahler, S. Todd Barlow, Brian A. Stone, and Ravinder S. Bhogal. 1997. The Persona Effect: Affective Impact of Animated Pedagogical Agents. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. ACM, 359–366.

[36] Ivan Leudar, Alan Costall, and Dave Francis. 2004. Theory of Mind (A Critical Assessment). *Theory & Psychology* 14, 5 (2004), 571–578.

[37] Michael Lewis. 2008. Self-Conscious Emotions: Embarrassment, Pride, Shame, and Guilt. In *Handbook of Emotions*, Michael Lewis, Jeannette M. Haviland-Jones, and Lisa Feldmann Barrett (Eds.). New York: The Guilford Press, 742–756.

[38] Christine L. Lisetti and Fatma Nasoz. 2002. MAUI: A Multimodal Affective User Interface. In *Proceedings of the tenth ACM international conference on Multimedia.* ACM, New York, NY, 161–170.

[39] Stacy Marsella and Jonathan Gratch. 2014. Computationally Modeling Human Emotion. *Commun. ACM* 57, 12 (2014), 56–67.

[40] Stacy C. Marsella and Jonathan Gratch. 2009. EMA: A process model of appraisal dynamics. *Cognitive Systems Research* 10, 1 (2009), 70–90.

[41] Stacy C. Marsella, Jonathan Gratch, and Paola Petta. 2010. Computational Models of Emotion. In *Blueprint for Affective Computing (A Sourcebook)*, Klaus R. Scherer, Tanja Bänzinger, and Etienne B. Roesch (Eds.). Oxford University Press, Oxford, 21–41.

[42] Jörg Merten. 1996. *Affekte und die Regulation nonverbalen, interaktiven Verhaltens: strukturelle Aspekte des mimisch-affektiven Verhaltens und die Integration von Affekten in Regulationsmodelle.* Lang.

[43] Jörg Merten. 2003. Context-analysis of facial-affective behavior in clinical populations. *The Human Face. Measurment and Meaning* (2003), 131–147.

[44] Agnes Moors, Phoebe C. Ellsworth, Klaus R. Scherer, and Nico H. Frijda. 2013. Appraisal Theories of Emotion: State of the Art and Future Development. *Emotion Review* 5, 2 (April 2013), 119–124.

[45] Marcello Mortillaro, Ben Meuleman, and Klaus R. Scherer. 2012. Advocating a Componential Appraisal Model to Guide Emotion Recognition. *International Journal of Synthetic Emotions (IJSE)* 3, 1 (2012), 18–32.

[46] Ulrich Moser. 2009. *Theorie der Abwehrprozesse (Die mentale Organisation psychischer Störungen).* Brandes & Apsel Frankfurt a. M.

[47] Ulrich Moser and Ilka von Zeppelin. 2005. *Psychische Mikrowelten - Neuere Aufsätze.* Vandenhoeck & Ruprecht, Chapter Die Entwicklung des Affektsystems, 161–221.

[48] Ulrich Moser, Ilka von Zeppelin, Rolf Pfeiffer, and Werner Schneider. 1991. *Cognitive - Affective Processes - New Ways of Psychoanalytic Modeling.* Springer, Berlin, Heidelberg.

[49] Ulrich Moser and Ilka von Zeppelin. 1996. Die Entwicklung des Affektsystems. *Psyche - Zeitschrift für Psychoanalyse und ihre Anwendungen* 50, 1 (1996), 32–84.

[50] Kevin P. Murphy. 2002. Dynamic Bayesian Networks. *Probabilistic Graphical Models, M. Jordan* 7 (2002).

[51] Donald L. Nathanson. 1994. *Shame and Pride: Affect, Sex, and the Birth of the Self.* WW Norton & Company.

[52] Andrew Ortony, Gerald L. Clore, and Allan Collins. 1988. *The Cognitive Structure of Emotions.* Cambridge University Press, Cambridge, MA.

[53] Brian Parkinson and Antony S. R. Manstead. 2015. Current Emotion Research in Social Psychology: Thinking About Emotions and Other People. *Emotion Review* (2015), 1754073915590624.

[54] Rolf Pfeifer. 1988. Artificial Intelligence Models of Emotion. In *Cognitive perspectives on emotion and motivation.* Springer, 287–320.

[55] Rosalind W. Picard. 1997. *Affective Computing.* MIT Press, Cambridge, MA.

[56] David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1 (1978), 515–526. Issue 04.

[57] Anand S. Rao and Michael P. Georgeff. 1995. BDI agents: From theory to practice.. In *ICMAS*, Vol. 95. 312–319.

[58] Sérgio H. Rodrigues, Samuel F. Mascarenhas, João Dias, and Ana Paiva. 2009. "I can feel it too!": Emergent Empathic Reactions Between Synthetic Characters. In *In 3rd Proceedings of International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009.* IEEE, 1–7.

[59] Luis-Felipe Rodríguez and Félix Ramos. 2014. Development of computational models of emotions for autonomous agents: a review. *Cognitive Computation* 6, 3 (2014), 351–375.

[60] Thomas J. Scheff and Suzanne M. Retzinger. 2000. Shame as the Master Emotion of Everyday Life. *Journal of Mundane Behavior* 1, 3 (2000), 303–324.

[61] Frank Schwab. 2000. *Affektchoreographien. Eine evolutionspsychologische Analyse von Grundformen mimisch-affektiver Interaktionsmuster.* Ph.D. Dissertation. Dissertation am FB 5.3 Empirische Humanwissenschaften der Universität des Saarlandes.

[62] Craig A. Smith and Phoebe C. Ellsworth. 1985. Patterns of cognitive appraisal in emotion. *Journal of personality and social psychology* 48, 4 (1985), 813.

[63] William R. Swartout, Jonathan Gratch, Randall W. Hill Jr., Eduard Hovy, Stacy Marsella, Jeff Rickel, David Traum, et al. 2006. Toward Virtual Humans. *AI Magazine* 27, 2 (2006), 96.

[64] Maya Tamir. 2011. The Maturing Field of Emotion Regulation. *Emotion Review* 3, 1 (2011), 3–7.

[65] Silvan S. Tomkins. 1984. Affect theory. *Approaches to emotion* 163 (1984), 195.

[66] Michel Valstar, Tobias Baur, Angelo Cafaro, Alexandru Ghitulescu, Blaise Potard, Johannes Wagner, Elisabeth André, Laurent Durieu, Matthew Aylett, Soumia Dermouche, et al. 2016. Ask Alice: An Artificial Retrieval of Information Agent. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction.* ACM, New York, NY, USA, 419–420.

[67] Susanne Van Mulken, Elisabeth André, and Jochen Müller. 1998. The Persona Effect: How Substantial Is It? *People and computers XIII: Proceedings of HCI* 98 (1998), 53–66.

[68] Vinobar Vinayagamoorthy, Marco Gillies, Anthony Steed, Emmanuel Tanguy, Xueni Pan, Céline Loscos, and Mel Slater. 2006. Building Expression into Virtual Characters. In *STAR Proceedings of Eurographics 2006*, Brian Wyvill and Alexander Wilkie (Eds.). Eurographics Association, Vienna, Austria, 21–61.

[69] Thurid Vogt and Elisabeth André. 2005. Comparing Feature Sets for Acted and Spontaneous Speech in View of Automatic Emotion Recognition. In *ICME.* IEEE, 474–477.

[70] Thurid Vogt, Elisabeth André, and Nikolaus Bee. [n. d.]. EmoVoice - A Framework for Online Recognition of Emotionsfrom Voice. In *Perception in Multimodal Dialogue Systems, 4th IEEE Tutorialand Research Workshop on Perception and Interactive Technologiesfor Speech-Based Systems, Kloster Irsee, Germany* (2008) *(Lecture Notes in Computer Science).* Springer, 188–199.

[71] Christian von Scheve. 2010. Die emotionale Struktur sozialer Interaktion: Emotionsexpression und soziale Ordnungsbildung/The Emotional Structure of Social Interaction: The Expression of Emotion and the Emergence of Social Order. *Zeitschrift für Soziologie* (2010), 346–362.

[72] Johannes Wagner, Tobias Baur, Yue Zhang, Michel F. Valstar, Björn Schuller, and Elisabeth André. 2018. Applying Cooperative Machine Learning to Speed Up the Annotation of Social Signals in Large Multi-modal Corpora. *arXiv preprint arXiv:1802.02565* (2018).

[73] Johannes Wagner, Florian Lingenfelser, Tobias Baur, Ionut Damian, Felix Kistler, and Elisabeth André. 2013. The Social Signal Interpretation (SSI) Framework: Multimodal Signal Processing and Recognition in Real-Time. In *Proceedings of the 21st ACM international conference on Multimedia.* ACM, 831–834.

[74] Yorick Wilks (Ed.). 2010. *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues.* Vol. 8. John Benjamins Publishing.

[75] Martin Wöllmer, Marc Al-Hames, Florian Eyben, Björn Schuller, and Gerhard Rigoll. 2009. A multidimensional dynamic time warping algorithm for efficient multimodal fusion of asynchronous data streams. *Neurocomputing* 73, 1 (2009), 366–380.

[76] Martin Wöllmer, Moritz Kaiser, Florian Eyben, Björn Schuller, and Gerhard Rigoll. 2013. LSTM-Modeling of Continuous Emotions in an Audiovisual Affect Recognition Framework. *Image and Vision Computing* 31, 2 (2013), 153–163.

[77] Atef Ben Youssef, Nicolas Sabouret, and Sylvain Caillou. 2014. Subjective Evaluation of a BDI-based Theory of Mind model. In *Workshop on Affect, Compagnon Artificiel, Interaction (WACAI)*. 120–125.

[78] Zhihong Zeng, Jilin Tu, Brian M. Pianfetti, and Thomas S. Huang. 2008. Audio–Visual Affective Expression Recognition Through Multistream Fused HMM. *IEEE Transactions on Multimedia* 10, 4 (2008), 570–577.