Cognition and Emotion in Autonomous Cars

Sven Kraus, Matthias Althoff, Bernd Heißing and Martin Buss

Abstract— In order to increase the acceptance of vehicles that drive (partly) autonomously, it seems advantageous that the driving style of autonomous cars is human-like. Furthermore, the acceptance of autonomous cars is believed to be increased when their actions and current state becomes more transparent to the passengers. These aspects can be tackled by implementing emotions to an autonomous car which allows a situation specific adaption of driving functions and vehicle dynamics. In this work, the emotions are modeled by the cognitive appraisal model as described by Ortony, Clore and Collins. The approach is presented for a traffic scene where two lanes are merged into one.

I. INTRODUCTION

Current research on autonomous cars focuses on accomplishing increasing complex driving tasks. Although remarkable progress has been made (which could be observed e.g. at the DARPA Urban Challenge), passengers would not feel comfortable in current autonomous cars. The reasons are mainly twofold: the autonomous cars drive robotic-like when following planned trajectories which often feels unnatural compared to human driving. The second reason is, that the passengers have no idea if the autonomous vehicle recognizes and evaluates traffic situation correctly or if a critical maneuver has to be performed. These two shortcomings can be diminished when implementing emotions into a car. This is because the emotional state of the car allows to adapt the vehicle dynamics to the current emotion such as joy, resentment or fear and furthermore can make the driving decisions much more transparent to the passengers in a natural way. The generation of the emotions is performed according the cognitive appraisal model of emotions by Ortony, Clore and Collins (OCC) [1]. The mentioned authors specifically developed their theoretical approach with a view to computational implementation. They postulate that emotions emerge as a result of cognitions and interpretations and therefore focus on the cognitive elicitors of emotions. A first adaption of this model to the context of autonomous driving has been presented in [2]. Other works that consider emotions in autonomous vehicles, but with another model for emotions, are e.g. [3], [4]. Further related work has been published dealing with the modeling of emotional agents, e.g. [5], [6]. Applications of emotion models can be found in the field of Human Machine Interaction like in embodied characters [7], as well as in decision making modules [8] and machine vision applications [9].

The presented work differs from the above mentioned by using emotion models in the entire perception-cognition-action loop of a real autonomous vehicle in order to implement a human-like situation adaptive driving style. This is done by adjusting the controllers of the autonomous driver by means of just one central situation assessment instance – the emotion model. The paper especially presents an approach to determine a specific emotion value given by a particular traffic scenario.

The paper is organized as follows: In Section II an introduction into the cognitive appraisal model according to [1] and its adaptations to the requirements of the application area *traffic* will be given. Section III describes the generation of the intensity value *crash probability* within this model that contributes to certain emotions. Finally, Section IV shows a numerical example. The approach presented herein towards an emotionally driving car has been carried out within the *Cognitive Automobiles* research project [10].

II. EMOTION MODEL

Autonomously operating automobiles have already been built at the beginning of the 50s of the last century. However, no design has been proposed yet, which provides autonomous cars with the capability of a deep (cognitive) understanding of their environment, which is required for situation aware driving. Marvin Minsky [11] already noted in 1988:

> 'The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions.'

Emotions determine our actions and the manner of the execution of those. They arise as a result of certain cognitions and interpretations of external influences from the environment, and their entity defines the inner state of a person or system.

In general, the selected emotions for artificial or technical systems are from a subset of the human emotions (such as fear, joy,...). Certain sensors are assigned to these inner states, whose data influence the emotions. Based on the current emotion values, decisions and behaviors can be derived. The advantage of this approach lies in the abstraction of the situation, which is complex due to the large number of sensor data. The data are aggregated into emotions according to specific and application adjusted rules and hence into higher-level information.

A. General Structure of the Model

The proposed emotion model, derived from [1], represents emotions as valenced reactions to certain perspectives of the world – in the presented application this perspectives are restricted to the traffic scenario. Emotions are distinguished between ones that elicit as reaction to the consequences of events, to the actions of agents or to the aspects of objects. Thus, three basic emotion classes arise: being *pleased* vs. *displeased* (reactions to events), *approving* vs. *disapproving*
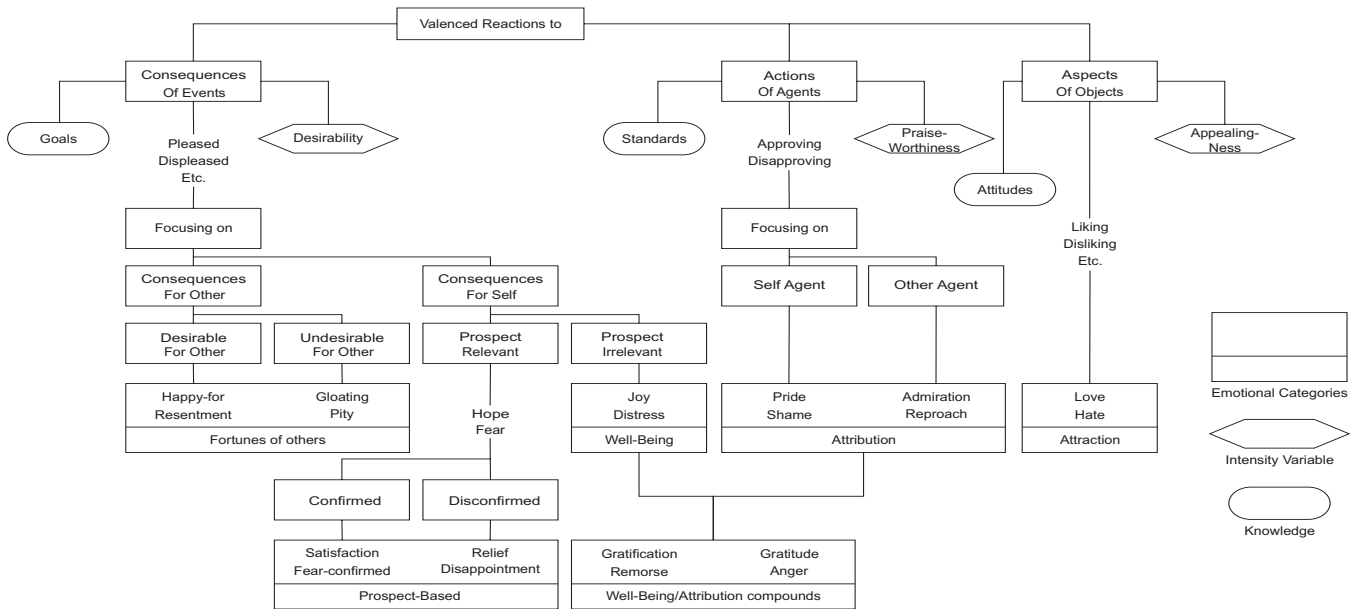
Fig. 1. The original OCC Model [1].

(reactions to agents) and *liking* vs. *disliking* (reactions to objects). In this use case events are considered to be any relevant traffic constellation disregarding the cause of their origin. Agents can be any kind of traffic participant, like bikes, motorcycles, passenger cars or trucks and their actions might be following, approaching, passing, overtaking, driving with constant speed or distance, lane changing and so on. Objects are defined as static elements of the traffic scenario, e.g. the road itself or the lane boundaries. Altogether, 22 emotion types are specified which can be differentiated into distinct groups of emotional categories (see Fig. 1).

B. Intensity Variables

One of the main questions is what determines the intensity of emotions. Ortony, Clore and Collins introduce a number of variables in order to answer this question. They distinguish between *Global*, *Local* and *Central Intensity Variables*. Intensity variables belonging to the first group, affect the intensity of all emotions. Among these are the variables *sense of reality*, *proximity*, *unexpectedness* and *arousal*. Intensity variables belonging to the second group, merely refer to emotions classified in certain emotional categories. These are *likelihood*, *effort*, *realization*, *desirability for other*, *deservingness*, *liking*, *strength of cognitive unit*, *expectation-deviation* and *familiarity*. For more detailed information we refer to [1], except for the computation of *likelihood* that will be explained in detail in Section III.

The appraisal of an emotion-inducing situation is mainly influenced by three *Central Intensity Variables*, namely *desirability*, *praiseworthiness* and *appealingness*. They affect the event-based, the agent-based and the object-based emotions. The latter are evaluated by the degree of liking certain aspects of the object depending on one's attitudes. Regarding, for example the object road, certain attitudes could be:

- the more drivable lanes, the better (because of lower traffic density on condition of constant number of traffic participants)
- the wider the lane, the better (because of higher lateral safety distances)
- the lower the curvature of the road, the better (because of lower lateral acceleration)

Actions of agents are assessed with respect to a certain catalog of universal traffic standards and norms. Depending on how much the action of an agent conforms to these standards, the higher the praiseworthiness of this action is judged. Transferred to the traffic scenario, such standards could be:

- formal: observing traffic rules like speed limits, safety distances, right of way,...
- informal: thoughtful driving, e.g. permit merging in

The consequences of events are considered as desirable or not desirable in terms of one's own goals and desires. These are structured in complex hierarchical networks considering different kinds of goals and subgoals with inherited dependences between them.

For a certain traffic situation main goals and subgoals can be identified according to Fig. 2. The connections between these are labeled by abbreviations expressing the kind of influence, that the subgoals exert on the higher goals. More precisely, the degree of sufficiency or necessity for achieving the higher goals. The abbreviations stand for **S**ufficient, **N**ecessary, **I**nhibitory and **F**acilitative. The facilitating links descend from goals, that when achieved, increase the probability of attaining higher goals, even though not guaranteeing it. In the example considered in this work, the main goals are *Accident-free Driving* and *Time Efficient Driving*, whereas the latter is mainly mentioned here to show, that subgoals like *Avoidance of Close Gaps* may exert in different ways

on several higher goals. If *Accident-free Driving* is seen as the crucial goal, e.g. in the situation of an ending lane before a road works occurs (example will be discussed in Section IV in more detail) and considerations are restricted to this goal then it will be influenced by three subgoals : *Keeping Safety Distances, Trajectory with low/zero Crash Probability* and *(Maladjusted) High Speed*. The goal *Keeping Safety Distances* affects the main goal in a facilitative and sufficient way, since going below the safety distances not automatically leads to an accident, but the risk of a collision is significantly increased. The choice of a *Trajectory with low/zero Crash Probability* is essential, however, and thus a necessary goal. *(Maladjusted) High Speed* - besides violating safety distances - is counted among the main causes of accidents on roads worldwide and is therefore an inhibitory goal regarding to *Accident-free Driving*. Conversely, the *(Maladjusted) High Speed* can be identified as a facilitative goal regarding higher objectives as *Time Efficient Driving* and so has a justification as a subgoal. Furthermore the identified subgoals can be clearly subdivided into tasks of longitudinal and lateral control in the autonomous driving mode. Therefore active influence on the experience of certain emotions becomes possible by the adaptation of the controllers. If, for example, a car cuts into our lane in front of us too close and the whole situation elicits a strong emotion of fear, then the control parameters of the distance controller are tightened and so lead to a more dynamic reaction that helps to defuse the situation faster.
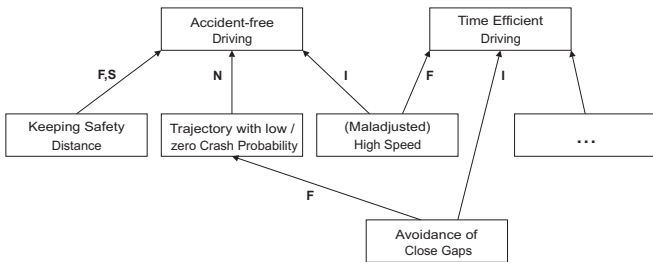

Fig. 2.   Goals in certain traffic situations

As an objective evaluation criteria of consequences of events with respect to the subgoal *Keeping Safety Distances*, the time gap computed of the relative distance and the actual driving speed (e.g. on motorways: value of the safety distance equates half the value of the speedometer, i.e. 1.8 s Time Gap) can be used. For the evaluation of *Trajectory with low/zero Crash Probability* an appropriate prediction of the situation with probabilistic reachable sets (according to following Section III) is carried out and the subgoal *(Maladjusted) High Speed* can be determined by a variable we call *TTB* (Time to Brake-Reaction). This is the time you have left before full brake deceleration is needed in order to just prevent a collision. It is calculated according to formula 1, where TTC is the *Time-to-Collision*, v_{rel} is the relative velocity and $a_{max(dec)}$ is the maximum deceleration.

$$TTB = TTC - 0.5 \cdot v_{rel}/a_{max(dec)} \qquad (1)$$

In this section, the cognitive appraisal model of Ortony, Clore and Collins and its adaptations to the use case traffic were presented. Thus, the possibility to derive emotions from traffic situations is given. Emotions are determined by certain computed variables that reflect valenced reactions to the construal of the actual traffic situation and that are suitable for activating certain emotions. So, in a concrete example of a fear-eliciting situation, first, the situation has to be evaluated in terms of the desirability of the consequences. If the event is displeased with respect to one's goals and the focus is on the consequences for oneself and the prospect of the event is relevant, then the next step would be to evaluate the local intensity variable *likelihood*. This will be done in the following section.

III. SAFETY ASSESSMENT

In principle, the safety assessment of the autonomous vehicle works like an artificial co-pilot that warns the vehicle when a dangerous situation is approaching. In order to determine the threat level of a situation, the following information has to be gathered by the scene interpretation module of the vehicle:

- the geometric description of the relevant road sections,
- the position and geometry of static obstacles,
- the position, velocity and classification of dynamic obstacles.

Additionally, the safety verification module requires the planned trajectory of the autonomous car, which is available within the vehicle anyway. Static obstacles are a special case of dynamic obstacles with zero velocity and thus the further discussion focuses on dynamic obstacles. The classification of dynamic obstacles ($\hat{=}$ other traffic participants) is understood as the grouping into *cars, trucks, motorbikes, bicycles* and *pedestrians*.

Measurements are subject to uncertainties which cannot be neglected. Thus, in this work, the initial state (position and velocity) of the traffic participants is modeled by a probability distribution. Based on a dynamic model of the traffic participants, the future probability distribution of positions and velocities is computed, where the acceleration of the traffic participants is also specified by a probability distribution. The probabilistic information about future positions of traffic participants allows to compute the probability of a crash when the autonomous car follows its planned trajectory. This is exemplarily shown for a simple traffic scene in Fig. 3 where the autonomous vehicle plans to turn into another street when another car is approaching. For the time interval $\tau_1 = [0, t_1]$, no crash can occur, as the stochastic reachable sets do not intersect, whereas a crash may occur in the next time interval $\tau_2 = [t_1, t_2]$.

A. Models of Traffic Participants

The paths of other traffic participants are chosen as the centerline of drivable lanes. The lateral deviation along these paths is modeled by a piecewise constant probability distribution $f(\delta)$ and δ is the deviation along the paths, see Fig 4. The probability distribution along the paths $f(s)$

is determined by the longitudinal dynamics of the traffic participants for the position s and the velocity v:

$$\dot{s} = v, \quad \dot{v} = \begin{cases} c_1 \cdot (1 - (v/c_2)^2) \cdot u, & u > 0 \\ c_1 \cdot u, & u \leq 0 \\ 0, & v \leq 0, v \geq v^{max} \end{cases} . \quad (2)$$

The constant c_1 models the maximum possible acceleration due to tire friction and c_2 the top speed - these constants are chosen according to the specific properties of the different classes of traffic participants. The acceleration input u is normalized and varies from $[-1, 1]$ covering all input commands from full braking to full acceleration. The constraint $\dot{v} = 0$, for $v \leq 0, v \geq v^{max}$ restricts backward driving and speeding (where v^{max} can be chosen a little higher than the official speed limit). Note that the suggested longitudinal dynamics model can be exchanged by other models without compromising the approach.

The lateral and longitudinal probability distribution can be combined to $f(s, \delta) = f(s) \cdot f(\delta)$ (see also Fig. 4 for an exemplary probability distribution of a vehicle on path 2) as independence is assumed. For unstructured scenarios, such as parking lots, the approach in [12] is suggested which uses the same mathematical principles as introduced herein.

B. Computing the Longitudinal Probability Distribution Using Markov Chains

The probability distribution $f(s)$ along the paths of other traffic participants is efficiently computed by abstracting the original system dynamics in (2) by Markov chains, see [13], [14]. This approach, which was previously proposed by the authors, is introduced in a concise fashion in the following. Markov chains are stochastic automata with discrete states $z \in \mathbb{N}^+$, which are not exactly computed, such that one computes with the probability of state values i: $p_i = P(z = i)$ and p is the probability vector of the Markov chain. By definition, the probability vector evolves according a linear map of the previous time step:

$$p(t_{k+1}) = \Phi \, p(t_k) \quad (3)$$

where $t_k = k \cdot T$ and $k \in \mathbb{N}^+$ is the time step and $T \in \mathbb{R}^+$ is the time step size. The probability values $\Phi_{ji} = P(z(t_{k+1}) = j | z(t_k) = i)$ contain the probability values that the state value changes from i to j after one time step. In order to abstract the continuous vehicle dynamics by a
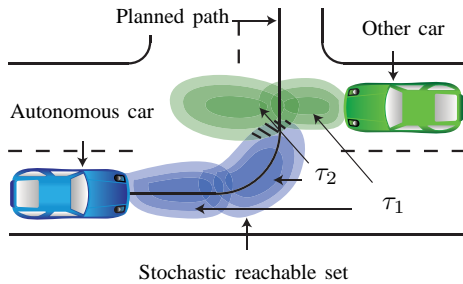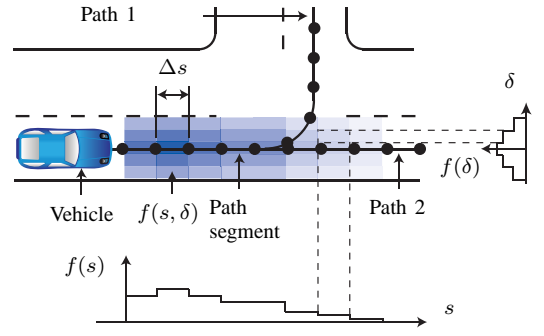

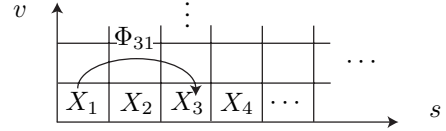Fig. 4. Probability distribution of the position along a path.

Fig. 5. Discretization of the state space.

Markov chain, two subtasks have to be dealt with: Relate the discrete states of the Markov chain to the continuous variables of position and velocity and compute the transition probabilities in Φ.

1) Discretization of the State and Input Space: The discrete states of the Markov chain are defined as state space regions of the previously continuous state space, spanned by the position and velocity of the vehicle, see Fig. 5. In order to end up with a finite number of discrete states, the state space is discretized for a subset $X \subset \mathbb{R}^2$ into rectangular cells X_i of equal size. Besides the state space, the input interval $u \in [-1, 1]$ is also discretized in equidistant intervals U^α and α is the value of the discrete input which is denoted y. In order to distinguish indices of state space cells from indices of input intervals, input indices are superscripted and Greek.

2) Transition Probabilities of the Markov Chains: The transition probabilities for different input intervals α are denoted $\Phi_{ji}^\alpha(T)$ and are defined as $\Phi_{ji}^\alpha(T) = P(z(t_{k+1}) = j | z(t_k) = i, y([t_k, t_{k+1}]) = \alpha)$ and $y([t_k, t_{k+1}])$ denotes the discrete input for the time interval $t \in [t_k, t_{k+1}]$. The transition probability from cell i to another cell j is based on the volume ratio of the reachable set R_i^α starting in cell i under input α that intersects with cell j (see Fig. 6(a)):

$$\Phi_{ji}^\alpha(T) = \frac{V(R_i^\alpha(T) \cap X_j)}{V(R_i^\alpha(T))}, \quad (4)$$

where $V()$ is an operator returning the volume. For further information on the computation of reachable sets, the reader is referred to [15]. In an analogous way, one can compute the transition probabilities $\Phi_{ji}^\alpha([0, T])$ for the time interval $t \in [0, T]$ by substituting $R_i^\alpha(T)$ with $R_i^\alpha([0, T])$ which is the reachable set for $t \in [0, T]$. The reachable set of $R_i^\alpha([0, T])$ is exemplarily shown for the vehicle dynamics (2) in Fig. 6(a). The corresponding stochastic reachable cells are illustrated in Fig. 6(b), in which the reachable set of Fig. 6(a) is indicated. A transition to a cell is the more likely,

Fig. 3. Stochastic reachable sets of traffic participants.

the darker the color of the cell is. Note, that the transition probabilities in $\Phi_{ji}^{\alpha}(T)$ and $\Phi_{ji}^{\alpha}([0,T])$ are computed offline, such that computationally expensive operations are performed beforehand.

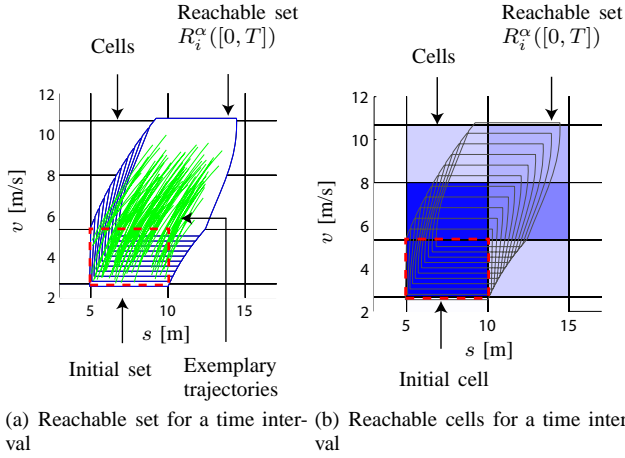(a) Reachable set for a time interval (b) Reachable cells for a time interval

Fig. 6. Reachable set of the original system and the corresponding stochastic reachable set of the Markov-chain.

C. Computing Stochastic Reachable Sets using Markov Chains

The computed transition probabilities $\Phi_{ji}^{\alpha}(T)$ and $\Phi_{ji}^{\alpha}([0,T])$ allow to compute the probability distribution of the next time step and time interval as shown in (3):

$$
\begin{aligned}
p(t_{k+1}) &= \Phi^{\alpha}(T)p(t_k), \\
p([t_k, t_{k+1}]) &= \Phi^{\alpha}([0,T])p(t_k).
\end{aligned}
\tag{5}
$$

The above computation in (5) can only be applied if the input α is known for each time interval. If this is not the case, one can define the joint probability $p_i^{\alpha} = P(z = i, y = \alpha)$ and compute the probability distributions as

$$
\begin{aligned}
p^{\alpha}(t_{k+1}) &= \Phi^{\alpha}(T)p^{\alpha}(t_k), \\
p^{\alpha}([t_k, t_{k+1}]) &= \Phi^{\alpha}([0,T])p^{\alpha}(t_k).
\end{aligned}
$$

In contrast to (5), the above computation has to be performed for all input values α, as the exact input is unknown. The change of the input probability distribution can be computed by a further Markov chain:

$$
p_i^{\beta}(t_k)' = \Gamma_i^{\beta\alpha} p_i^{\alpha}(t_k).
\tag{6}
$$

The prime indicates the instant change of the probability distribution at time t_k and Γ is the transition probability that the input value changes from $y = \alpha$ to $y = \beta$ when the state is $z = i$: $\Gamma_i^{\beta\alpha} = P(y' = \beta | y = \alpha, z = i)$. The transition matrices Γ_i can be learned by observation of traffic scenes or set by a combination of simulations and heuristics, where the latter is used in this work. A possible way of specifying Γ_i is given in [16].

IV. NUMERICAL EXAMPLE

The presented concept for generating emotions in an autonomous car is demonstrated for a traffic scenario where two lanes merge into one. For this reason, the autonomous car has to perform a lane change while respecting the vehicles on the neighboring lane. Furthermore, the vehicle has to consider that the location where the lanes merge into one is coming closer, such that an action becomes imminent. The probabilistic prediction of the traffic scene can be seen in Fig. 7 for a prediction that was computed at the time t_0. The autonomous car is indicated by an A and the vehicles B and C are the vehicles on the right lane, see Fig. 7. The initial speed of vehicle B (uniform probability distribution within 30 ± 2 m/s) is chosen greater than the one of vehicle C (uniform probability distribution within 20 ± 2 m/s), such that the gap for a lane change becomes closer over time. This is why the lane change is more risky when executed at the time point $t_3 = t_0 + 1.5$ sec as shown by the traffic prediction in Fig. 8. The traffic scene have been predicted for a time horizon of $t_f = 3$ sec (the plots do not show all predicted time intervals) which took 0.59 sec in Matlab on a 3.7 GHz single core PC, such that the prediction is about 5 times faster than real time.
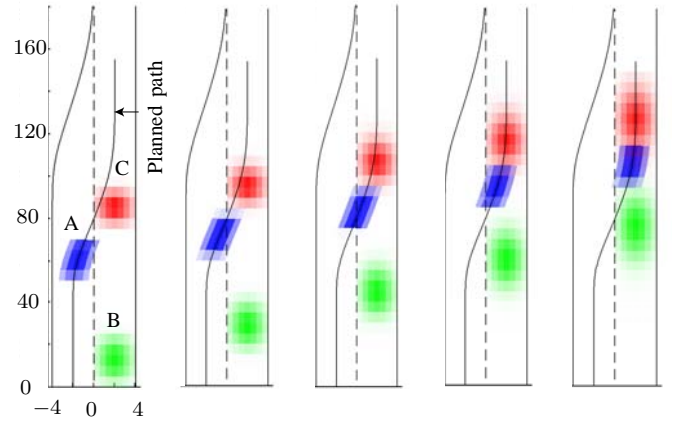
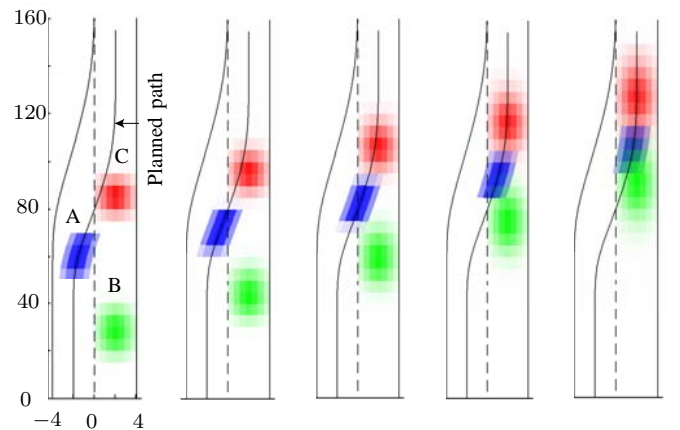Fig. 7. Stoch. reachable sets for the lane change scenario computed at t_0.

Fig. 8. Stoch. reachable sets for the lane change scenario computed at t_3.

Regarding the intensity variables introduced in Sec. II-B and their effects on the emotion type *fear*, the progress of this emotion value can be seen in Fig. 10. Considering the

consequences of the presented event we have to evaluate its desirability in terms of the goals according to Fig. 2. Thus, the closer the autonomous car A gets to the lane ending without considering a lane change, the lesser the time gap to the road construction, the higher the negative contribution to the goal *Keeping Safety Distance*. This equals for the progress of the objective evaluation criteria *Time to Brake-Reaction (TTB)*, that influences the goal *(Maladjusted) High Speed*. The less the TTB, the more it contributes to the maladjustment of the driven speed, which in turn inhibits to attain the main goal *Accident-free Driving*. Furthermore, with regard to the next step in emotion processing (following the branch of *Prospect-Based Emotions*) the progress of likelihood, formally represented as the maximum crash probability, is depicted in Fig. 9.
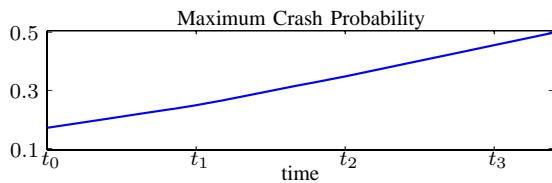

Fig. 9. Progress of the maximum crash probability

In summary, the intensity value of the emotion *fear* is a function of both the values of *desirability* and *likelihood* that may be additionally emphasized by the values of the global intensity variables that affect all emotions in general. Additionally, the progress of another emotion value – reproach – is shown in Fig. 10. The narrower the gap for car A gets to pull out into moving traffic, the higher the reproach of the behavior of car B is judged with regard to the informal standard of permitting to merge in. It has to be mentioned, that both the thresholds above which certain emotions just elicit and the maximum value of them first have to be evaluated on the basis of proband studies in the near future. So the depicted figure has to be considered as qualitative progress of the computed emotion values.
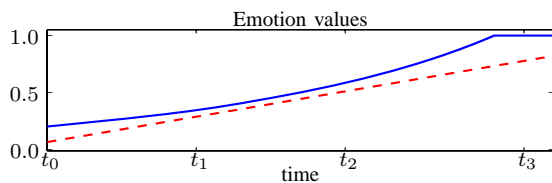

Fig. 10. Progress of emotion value fear (solid) and reproach (dashed)

V. Conclusion

In the context of a cognitive vehicle, the presented approach allows an emotional interpretation of the scene by the perception of the traffic area and its participants with their descriptive parameters. It is an approach that bundles the multitude of environmental variables, which the system *Cognitive Automobiles* is exposed, by a smaller number of theoretically interpreted emotional states. Through this data reduction the manageability and clarity of traffic situations can be simplified. The overall concept offers a huge potential for applications of emotion models in the area of advanced driver assistance systems. So, for example, depending on the emotional state of the system, an adjusted driving style could be established by adaptive control strategies. Experiments with different kinds of predefined maneuvers (like lane changing, lane keeping etc.) from normal to fear-like are part of our actual investigations. It is also possible to select certain actions, in a way, that the emotional state of the vehicle is steered away from the current emotion towards a target emotion. Latter may, for example, be characterized by a low degree of risk-taking and thus contributes to road safety.

References

[1] A. Ortony, G. Clore, and A. Collins, *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.
[2] D. M. Reichardt, "Approaching driver models which integrate models of emotion and risk," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2008, pp. 234–239.
[3] B.-K. Quek, J. Ibanez-Guzman, and K.-W. Lim, "A survivability framework for the development of autonomous unmanned systems," in *In Proc. of Control, Automation, Robotics and Vision*, 2006, pp. 1–6.
[4] C. Ippolito and G. Pisanich, "Cognitive emotion layer architecture for intelligent uav planning, behavior and control," in *In Proc. of Aerospace Conference*, 2005, pp. 1–16.
[5] M. Turkia, "A computational model of affects," *CoRR*, vol. abs/0811.0123, 2008.
[6] D. Barteneva, N. Lau, and L. P. Reis, "A computational study on emotions and temperament in multi-agent systems," *CoRR*, vol. abs/0809.4784, 2008.
[7] C. Bartneck, "Integrating the occ model in embodied characters," in *Proc. of the Workshop on Virtual Conversational Characters: Applications, Methods, and Research Challenges*, Melbourne, 2002.
[8] K. Mahboub, "Simulation of emotional processes in decision making," Master's thesis, Le Havre University, 2006.
[9] P. R. D. Silva, M. Osano, A. Marasinghe, and A. P. Madurapperuma, "A computational model for recognizing emotion with intensity for machine vision applications," *IEICE - Trans. Inf. Syst.*, vol. E89-D, no. 7, pp. 2171–2179, 2006.
[10] C. Stiller, G. Färber, and S. Kammel, "Cooperative cognitive automobiles," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2007, pp. 215–220.
[11] M. Minsky, *Society of Mind*. Simon & Schuster, 1988.
[12] F. Rohrmüller, M. Althoff, D. Wollherr, and M. Buss, "Probabilistic mapping of dynamic obstacles using markov chains for replanning in dynamic environments," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008, pp. 2504–2510.
[13] M. Althoff, O. Stursberg, and M. Buss, "Online verification of cognitive car decisions," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2007, pp. 728–733.
[14] ——, "Stochastic reachable sets of interacting traffic participants," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2008, pp. 1086–1092.
[15] ——, "Reachability analysis of nonlinear systems with uncertain parameters using conservative linearization," in *Proc. of the 47th IEEE Conference on Decision and Control*, 2008, pp. 4042–4048.
[16] ——, "Safety assessment of driving behavior in multi-lane traffic for autonomous vehicles," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2009.