Technical Details of a Domain-independent Framework for Modeling Emotion

Jonathan Gratch
University of Southern California
Institute for Creative Technologies
13274 Fiji Way, Marina del Rey, CA 90292
gratch@ict.usc.edu

Stacy Marsella
University of Southern California
Information Sciences Institute
4676 Admiralty Way, Marina del Rey, CA 90292
marsella@isi.edu

ABSTRACT

This technical report elaborates on the technical details of the EMA model of emotional appraisal and coping. It should be seen as an appendix to the journal article on this topic (Gratch & Marsella, to appear)

1. INTRODUCTION

This article specifies details of how EMA represents and reasons about causal events, beliefs, desires, intentions, actions. EMA relies on a cognitive component that combines techniques form decision-theoretic planning with models of beliefs, desires and intentions. Syntactic features of the resulting representations are then appraised along a number of dimensions, resulting in a set of appraisal frames and emotional responses. This article provides details of this process that were excluded from (Gratch & Marsella, to appear)

2. Cognitive Antecedents of Appraisal

EMA relies on a variant of classical planning techniques to model the causal reasoning that underlies many appraisal variables. The plan reasoning and representations underlying EMA combines aspects of two separate families of AI planning systems. On the one hand, it builds on ideas for integrating planning and execution, exemplified by Ambros-Ingerson and Steel's IPEM system (Ambros-Ingerson & Steel, 1988) and Golden et al.'s X11 (Golden, Etzioni, & Weld, 1994) planner. On the other hand, it builds on plan adaptation strategies initially proposed by Hayes (1975) and significantly elaborated by (Kambhampati & Hendler, 1992). The resulting hybrid supports planning, execution and replanning for environments where actions have duration and the world may change in surprising ways.

Plans are constructed via constraint posting as in other classical planners such as SNLP (McAllester & Rosenblitt, 1991). Constraints are added in response to perceived problems or ambiguities in the current plan. For example, if an action in the plan has an unestablished precondition, this threat may be resolved by identifying an existing action that establishes the effect (*simple-establishment*) or introducing a new action (*step-addition*). Either "fix" introduces new constraints to the current plan network. Simple-establishment asserts a *protection constraint* that protect the effect from the moment it is created until it is used by the precondition, and binding constraints that ensure the effect unifies with the open-precondition. Step-addition posts a constraint to include the new action in addition to the constraints posted by simple-establishment. Plan generation can be viewed as a sequential decision process where the planner repeated analyzes the current plan network and chooses on of a set of possible fixes.

The planner adopts IPEM's basic approach to integrating execution with this basic plan generation scheme. Besides maintaining a plan network, the planner maintains a declarative representation of the perceived state of the world or *current world description* (CWD). The CWD allows the planner to monitor the execution of tasks and detect any surprising changes in the environment. The planner also incorporates a set of execution "fixes" which may be interleaved with plan generation fixes. The planner may initiate tasks whose preconditions unify with the CWD (and are not preceded by any uninitiated tasks), terminate tasks who's effects appear in the CWD, and fail tasks if some pre-specified criterion is satisfied. As the CWD reflects the perceived state of the world, it may change in ways not predicted by the current plan network. For example, some external process modifying the environment is detected by changes to the CWD not predicted by the current set of executing tasks. These changes may provide opportunities (as when an unsatisfied precondition is unexpectedly observed in the world). They may also threaten constraints in the plan network, forcing the planner to modify the task network to resolve them.

In the Mission Rehearsal Exercise (MRE) application (Rickel et al., 2002), the planner incorporates two simplifications not usually adopted by classical planners. Rather than generating plans from scratch, steps are added from a pre-generated library of possible plans with limited branching. Further, the domain theory must be propositional (predicates cannot have free variables). These two simplifications greatly reduce the computational complexity of planning in the MRE but could be relaxed in other applications of EMA.

2.1 Representations

The planner must be able to represent its beliefs about the current state of the world, its goals, its plans, and possibly its beliefs about the plans of other agents (as discussed later). These representations are currently built on top of a general reasoning system called Soar (Newell, 1990) that uses a simple attribute value knowledge representation and rule-based reasoning (although the planning algorithm is not Soar-specific). Plans are maintained in a data structure called a *plan network.* The plan network consists of a set of tasks and a set of constraints over tasks or their sub-components.

2.1.1 Action representation

Tasks (sometimes referred to as operators, actions, or steps) represent the basic activities that an agent can perform in the world. Tasks are represented using the STRIPS formalism (with slight modification). Tasks have preconditions and effects, an execution state, and some other more specialized fields that will be described below. An example of a task definition is:

```
defTask 4th-sqd-recon-fwd {
   :agent 4sldr
   :destination celic
   :event recon
   :path route
   :pre {4th-sqd-at-aa}
   :add {4th-sqd-at-celic secure-route(0.75)}
   :del {4th-sqd-at-aa}
}
```

This definition denotes the act of the 4th Squad performing a reconnaissance forward along our route to the destination “Celic.” This can occur if the squad is at the assembly area (4th-sqd-at-aa) and results with the squad being at “Celic” and the route being secure with some probability (75%). The terms :agent, :destination, :event, and :path are utilized by the natural language processing module and do not impact the behavior of the planner.

Preconditions describe what must (necessarily) hold to successfully execute the task. The preconditions of a task correspond to a logical conjunction of predicates.

*Effects***:** The consequences of executing a task are described by an “add” list and “del” list of predicates. Informally, these lists describe facts that are made true or false as a consequence of executing the task. As tasks have duration, we must reconsider the standard STRIPS semantics that assumes that effects occur instantaneously (i.e., tasks define a discrete transition two quiescent states). Under the STRIPS assumption, add and delete lists correspond to logical conjunctive expression describing the difference between these two states. In the real world, effects do not happen instantaneously, nor do they occur simultaneously. Another complication is that, though real-world actions sometimes fail. These considerations violate the standard semantics of action effects.

As a consequence of these factors, we provide an alternative (informal) semantics for effects. The add list is a set of predicates that (individually) will (1) be satisfied by the CWD at some point during the execution of the task and (2) persist until some other activity in the world negates this fact. The delete list has the same semantics except that the predicates are implicitly negated. There is no closed world assumption.

A number of factors may prevent all of the effects of a task from being simultaneously realized. Tasks may fail causing only a subset of their effects to occur. Other simultaneously executing tasks may undo an effect during the task's execution. We allow multiple tasks to execute simultaneously but the representation language cannot generally express interactions between tasks (though Pednault illustrates how to represent some interactions in (Pednault, 1986)). If a task has different effects when other tasks are executing simultaneously, the planner won't properly predict these conditional outcomes (but it can react to them after the fact when the CWD changes in ways not predicted by the planner).

*Execution State***:** Each task instantiation in the plan network has an execution state attribute that describes its current execution status. Before a task executes it is in a *pending* state. After it has been initiated, its state transitions to *executing*, and after termination it becomes *executed.*

2.1.2 Constraint Representation

In addition to tasks, plans contain a number of constraints of various types. As mentioned above, CFOR is a constraint-posting planner. Planning is seen as a process of looking for possible violations of existing constraints (threat detection) and asserting new constraints to resolve possible violations (threat resolution). Constraint posting planners also perform some limited inference on constraints (or constraint propagation) and consistency checking (to look for possible constraint violations).

Ordering constraints: The planner can represent a partial ordering relationship between tasks. Following standard planning convention, an ordering constraint is a binary relation between tasks. Asserting *before(T1,T2)* means that task T1 occurs before task T2. Actually, things are a little more complicated because tasks have duration and are explicitly initiated and terminated. The constraint *before(T1,T2)* is interpreted to mean that task T1 will be initiated before task T2 is initiated (i.e., the start of T1 occurs before the start of T2). Ordering constraints are transitive: *before(T1,T2)* and *before(T2,T3)* implies *before(T1,T3)*. It is also assumed that plans are acyclic. Therefore, a set of ordering constraints containing a cycle is considered inconsistent.

Protection Constraints and Causal Links: Interval protection constraints (IPCs) assert that the truth value of some predicate must hold during the interval that occurs between two tasks. For example *IPC(T1, P, T2)* corresponds to the constraint that the predicate *P* must be true in any state occurring between the initiation of task T1 and the initiation of task T2. If some effect of some action violates an IPC (e.g, asserts NOT(P) in the interval between T1 and T2) the planner signals the IPC is threatened.

3. Probabilistic Reasoning and Coping Potential

Appraisal variables such as likelihood, unexpectedness and future expectancy involve some notion of probabilistic reasoning: How likely is this desirable outcome? How probable is this threat to my plans? In EMA, we focus on likelihood. To support the requirements of appraisal and coping, likelihood must be closely integrated with reasoning about plans and threats. To support coping, the assessed likelihood of outcomes also seems to be influenced by coping strategies (e.g., wishful thinking). A Bayesian statistical interpretation of likelihood is consistent with these requirements. The likelihood of events is characterized by a subjective probability that is "updated" by inference or observation. For example, an agent might attribute some *a priori* probability to the attainment of a goal. This probability can be subsequently refined by generating a specific plan to achieve the goal, essentially re-casting the probability of the goal in terms of the probability of more immediate subgoals. It is also clear that people used flawed probability models (Tversky & Kahneman, 1983). Rather than implementing full Bayesian inference, the current model uses a simpler approach to deriving these probabilities based on a strong independence assumption (all joint probabilities are modeled as the product of their constituent probabilities). This suffices as a first approximation and greatly simplifies a number of equations.

To assess coping potential, we must model not just the likelihood of goal attainment, but also the availability of alternative ways of achieving a goal. A number of theories of counterfactual reasoning point to the relationship between the "mutability" of a plan and the intensity of emotional responses that may result (Kahneman & Miller, 1986). For example, one might imagine two possible plans to achieve a goal, one with a high likelihood of goal attainment but which irrecoverably fails under certain circumstances, the other with less overall likelihood of success but which is easier to invent an alternative plan. One might imagine a complex model of counterfactual reasoning or probability distributions to capture this distinction. Here we introduce an additional parameter to capture the likelihood that a plan may be repaired.

3.1 Probability calculus

In EMA, the causal interpretation contains three classes of probabilities. *Achievement probabilities* represent the *a priori* probability that some goal or precondition can be achieved given its current truth-value in the current world description, denoted P_{ACH}(pre) where *pre* is the precondition of some action or a top-level goal. *Execution probabilities* represent the probability that an effect of an action will be achieved if the action is executed. Different effects may occur with different probability, denoted as P_{EX}(eff|I) where *eff* is the effect of an action and *I* denotes if the action is intended. *Repair probabilities* represent the probability that a plan can be repaired given that the current plan fails, denoted P_{REP}(pre) where *pre* is the precondition of some action or a top-level goal. If there is no existing plan, the repair probability defaults to the achievement probability for that precondition. This parameter represents a measure of how well the agent can cope with threats to the achievement of the effect.

The probability of actions, preconditions, and effects are determined by some simple rules acting on these probability values, as expressed in the current causal interpretation. The initial probability a top-level goal is its *a priori* achievement probability. This becomes refined as the planner refines its plans for the goal. The probability model propagates base probabilities through the interpretation using simple rules that key off of syntactic properties of plans. A (sub)goal is considered *established* if there is an establishment relation between the goal and some effect in the causal interpretation. This *establisher* is considered *threatened* if some other effect – called the *threat* – possibly undoes it before the (sub)goal is needed. An action may be pending or already initiated. The probability that a pending action will be executed depends on the probability its preconditions will be satisfied but also depend on whether or not the action is intended, as an unintended action is (presumably) less likely to be executed and the effect less likely to occur (Equation 2). EMA 's planning model assumes actions have duration so an action may be initiated and its effects observed somewhat later. An effect is *satisfied* as long as it is believed true in the current world description.

Since we have not addressed the problems of plan recognition or abduction, EMA currently uses an overly simplistic representation of the probability of past events. If an event is in the causal history, it is assumed that it occurred with certainty (probability of 1.0). Ideally, events and causal relations in the causal history would also be characterized by probabilities that would reflect one's belief in these historical interpretations.

Probability of an effect: P(eff)

IF State(action(eff)) = Pending THEN

$$P(\text{eff}) = P_{EX}(\text{eff}) * P(\text{intend}(\text{action}(\text{eff}))*\prod P(\text{precondition}(\text{action}(\text{eff})) \quad (1)$$

IF State(action(eff)) = Executing AND -satisfied(eff) THEN $P(\text{eff}) = P_{EX}(\text{eff})$ (2)

IF State(action(eff)) = (Executing OR Executed) AND satisfied(eff) THEN P(eff) = 1 (3)

IF State(action(eff)) = Executed AND -satisfied(eff) THEN P(eff)=0 (4)

Probability of a goal/precondition: Pr(goal)

IF -established(goal) THEN $P(\text{goal}) = P_{ACH}(\text{goal})$ (5)

IF established(goal) AND -threatened(goal) THEN P(goal) = P(establisher(goal)) (6)

IF established(goal) AND threatened(goal) THEN

$$P(\text{goal}) = P(\text{establisher}(\text{goal}))[1 - P(\text{threat}(\text{goal}))] + P_{REP}(\text{goal})P(\text{threat}(\text{goal})) \quad (7)$$

Equation 1 states that if an effect is associated with a pending action, the probability that it will be satisfied is its execution probability times the likelihood that the action can be initiated (which adopts the simplify assumption of precondition independence), times the likelihood that it is intended (as an unintended action is not likely to be executed, even if its preconditions are satisfied). If the action is executing, the probability depends on if the effect has been observed, while an effect that has not yet been observed after the action has terminated is assumed to have failed (probability equals zero). The probability of preconditions of actions depends on if they are established or not. Unestablished goals are assigned a probability equal to their achievement probability (Equation 5). Established and unthreatened goals will be established with a probability equal to the probability of the effect that establishes them (Equation 6). If a goal is threatened there are two possibilities. Either the threat will not occur (with probability 1-P(threat(goal)), or if it does occur, the plan may be repaired with some probability. Equation 7 summarizes this.

3.2 Extrinsic Utility

A number of appraisal theories emphasize that appraisal must account both for immediate rewards and more distant threats (Scherer, 2001). For example, a agent might conclude that flying to Chicago is desirable because it values being in Chicago as an end in itself. Sometimes, however, states gain importance because they are a means to an end. The agent might assign weight to being in Chicago because it is a step towards arriving in New York, the intrinsic goal. People can experience distress in response to subgoal violations as well as intrinsic goal violations, and a computation model of appraisal needs to facilitate such assessments by automatically deriving the intermediate (or "extrinsic") importance of these subgoals.

Following (Sloman, 1987) and (Beaudoin, 1995), EMA distinguishes between the intrinsic and extrinsic utility of states. A state's extrinsic utility relates to how it furthers other intrinsic goals. In their computational approach, Sloman and Beaudoin define extrinsic worth in terms of syntactic characteristics of the plan (e.g., the depth the goal falls in the plan hierarchy, the number of operators that could achieve the goal, etc.). In our view, this syntactic characterization amounts to a heuristic for assessing how much a subgoal's achievement contributes to the probability of attaining intrinsic goals. Rather, we explicitly define this contribution in terms of the change in the probability of intrinsic goal achievement: the utility of a subgoal is the sum of the intrinsic utility of goals it helps establish, weighted by how much its establishment adds to the probability each of these intrinsic goals will be achieved. Again, plan representations are key in this computation.

Extrinsic utility depends on two factors. First, we must identify all of the intrinsic goals impacted by a subgoal. Second, we must identify how much the subgoal impacts the attainment of each of these intrinsic goals. The set of impacted goals is simply the set of goals with intrinsic utility that are directly or indirectly connected to the subgoal via plans (in the transitive closure of the establishment relation). Computing the probabilistic contribution of the subgoal to each impacted goal has been studied in the decision-theoretic planning community. We adopt a much simpler (but not necessarily accurate) computation that exploits the independence of precondition probabilities assumed above. Given this strong independence assumption, the probability that a given goal will be achieved is simply the product of all of the "leaves" in the plan that achieves it (i.e., the set of preconditions in the transitive closure of the establishment relation). Thus, by dividing the current probability of an intrinsic goal by the current probability of some subgoal, we can separate the contribution of this subgoal from all other factors. The ratio defines the contribution of everything except the subgoal towards the intrinsic goals achievement. To assess the probabilistic contribution of this subgoal towards the intrinsic goal's achievement we simply measure the difference in probability of the intrinsic goal depending on if the subgoal were made true (subgoal probability equals one) vs. if the subgoal were not currently true (which corresponds to its achievement probability). The extrinsic utility of a subgoal s is defined as the sum of the utility of each impacted goal weighted by the change in probability that attaining the subgoal would have on it.[1]

[1] For efficiency reasons, EMA only computes extrinsic utility for appraisals assessed from the agent's own perspective. Only intrinsic utility values are used for appraisals of other agents.

$$Extrinsic\_Utility(s) = \sum_{g \in Impact(s)} Intrinsic\_Utility(g) \frac{P(g)[1 - P_{REP}(s)]}{P(s)} \quad (8)$$

The overall utility of the subgoal would be its extrinsic utility plus any intrinsic utility it might have. Qualitatively, this equation will give more utility to difficult to achieve subgoals (subgoals with low achievement probability). If a plan involves several preconditions, most of which are easy to achieve, only those with low achievement probability will have significant extrinsic utility. These can be viewed as the bottlenecks in the plan. Thus, only the achievement, or threats to the achievement of these subgoals will generate significant emotional responses. By relying on these achievement probabilities, this equation encodes a sort of conterfactual reasoning in the sense of (Kahneman & Miller, 1986), as only plan steps for which there are no available alternatives will generate significant emotional responses.

In the example, Dr. Tom does not attribute intrinsic utility to the state that morphine is approved. This state only inherits utility through its contribution towards the "downstream" states of making ending suffering and hastening death. Given the strong disutility of the later, Dr. Tom attributes negative utility to the treatment being approved, and will be distressed if this state is achieved.

4. Appraisal rules

Table 4 lists the set of rules used to map from features of the causal interpretation to individual appraisal frames and configurations of appraisal variables. Relevance rules examine every state fluent in the causal history and task network and, if the state has sufficient (dis)utility for the agent or if the state is believed to have sufficient intrinsic (dis)utility for other known agents, a frame is created to summarize the characteristics of the state. A given state may have multiple facilitators or inhibitors so, for each frame, a subframe is created to appraise each facilitation or inhibition event. Emotion instances are generated from individual subframes.

If a desired state has both an establisher and a threat, two separate frames represent this: one focusing just on the establisher and one just on the threat. This separation is an intentional design choice, allowing the agent to simultaneously "feel" both hope and fear about the same goal, possibly focusing more on one or the other depending on how natural language or other cognitive processes access the causal interpretation. The more conventional approach used in decision theory is to average these two factors into an overall expected utility, which tends to wash out such distinctions.

Appraisal rules are implemented as Soar elaboration rules. Soar implements elaboration rules via a justification-based truth maintenance system (Doyle, 1979). If the preconditions of the rule match elements of working memory, the effects of the rule are added to working memory. However, if one of the preconditions is subsequently retracted, the consequences are retracted as well. Thus, if after some planning an unestablished goal becomes established, the inhibition subframe and emotion instance related to the plan blockage will automatically retract and a new facilitation subframe and emotion instance will be created in its place.

Table 1: Appraisal Rules

$\forall$ state $\in$ causal_interpreation	IF |Utility$_{SELF}$(state)| > 1.0 THEN Create frame F: F:Relevant := True F:State := state; F:Perspective := Self F:Utility := Utility$_{SELF}$ (state)	(9)
	IF |IntrinsicUtility$_{OTHER}$(state)| > 1.0 THEN Create frame F: F:Relevant := True F:State := state; F:Perspective := Other F:Utility := Intrinsic_Utility$_{OTHER}$ (state)	(10)
$\forall$ F $\in$ Frames, $\forall$ S $\in$ F	IF established(F:State) THEN Create subframe F.S F.S:Type := facilitator; F,S:Cause := establisher(state) F.S:Likelihood := Pr(establisher(state)) F.S:Desirability := F:Utility F.S:Coping_Potential := Pr(establisher(state))	(11)
	IF established(F:State) AND threatened(F:State) THEN Create subframe F.S F.S:Type := inhibitor; F.S:Cause := threat(state) F.S:Likelihood := Pr(threat(state)) F.S:Desirability := 0.0 – F:Utility F.S:Coping_Potential := Pr$_{REP}$(state)	(12)
	IF unestablished(F:State) THEN Create subframe F.S F.S:Type := inhibitor; F.S:Cause := plan-blockage; F.S:Likelihood := Pr$_{ACH}$(state) F.S:Desirability := 0.0 – F:Utility F.S:Coping_potential := Pr$_{ACH}$(state);	(13)
	IF F.S:Desirability > 0 AND responsibility(F.S:Cause) = agent THEN F.S:Causal_attribution := praiseworthy F.S:Causal_agent := agent	(14)
	IF F.S:Desirability $<$ 0 AND responsibility(F.S:Cause) = agent THEN F.S:Causal_attribution := blameworthy F.S:Causal_agent := agent	(15)

Acknowledgements

This work was funded by the Department of the Army under contract DAAD 19-99-D-0046. Any opinions, findings, and conclusions expressed in this article are those of the authors and do not necessarily reflect the views of the Department of the Army.

References

Ambros-Ingerson, J., & Steel, S. (1988). *Integrating Planning, Execution and Monitoring.* Paper presented at the Seventh National Conference on Artificial Intelligence, St. Paul, MN.

Beaudoin, L. (1995). *Goal Processing in Autonomous Agents* (Ph.D Dissertation No. CSRP-95-2): University of Birmingham.

Doyle, J. (1979). A Truth Maintenance System. *Artificial Intelligence, 12*, 231-272.

Golden, K., Etzioni, O., & Weld, D. (1994). *Omnipotence without Omniscience: Efficient Sensor Management for Planning.* Paper presented at the National Conference on Artificial Intelligence, Seattle, WA.

Gratch, J., & Marsella, S. (to appear). A domain independent framework for modeling emotion. *Journal of Cognitive Systems Research*.

Kahneman, D., & Miller, D. (1986). Norm Theory: Comparing reality to its alternatives. *Psychological Review, 93*(2), 136-153.

Kambhampati, S., & Hendler, J. (1992). A validation-structure-based theory of plan modification and reuse. *Artificial Intelligence, 55*, 193-258.

McAllester, D., & Rosenblitt, D. (1991). *Systematic Nonlinear Planning.* Paper presented at the National Conference on Artificial Intelligence.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Pednault, E. P. (1986). Formulating Multiagent, Dynamic-world Problems in the Classical Planning Framework. In M. Georgeff & L. Lansky (Eds.), *Reasoning about Actions and Plans*: Morgan Kaufmann Publishers, Inc.

Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D., & Swartout, W. (2002). Toward a New Generation of Virtual Humans for Interactive Experiences. *IEEE Intelligent Systems, July/August,* 32-38.

Scherer, K. R. (2001). Appraisal Considered as a Process of Multilevel Sequential Checking. In K. R. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal Processes in Emotion: Theory, Methods, Research* (pp. 92-120): Oxford University Press.

Sloman, A. (1987). Motives, mechanisms and emotions. *Cognition and Emotion, 1*, 217-234.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgement. *Psychological Review, 90*(4), 293-315.