Österreichisches Forschungsinstitut für / Austrian Research Institute for / Artificial Intelligence

TR–2004–23

Stefan Rank, Paolo Petta, Robert Trappl

Features of Emotional Planning in Software Agents

• Freyung 6/6 • A-1010 Vienna • Austria •

• Phone: +43-1-5336112 •

• `mailto:sec@ofai.at` •

• `http://www.ofai.at/` •

Österreichisches Forschungsinstitut für / Austrian Research Institute for / Artificial Intelligence

TR–2004–23

Stefan Rank, Paolo Petta, Robert Trappl

Features of Emotional Planning in Software Agents

The Austrian Research Institute for Artificial Intelligence is supported by the Federal Ministry of Education, Science and Culture.

Citation: Rank S., Petta P., Trappl R. (2006): Features of Emotional Planning in Software Agents. In Della Riccia G., Dubois D., Kruse R., Lenz H.-J. (eds.): Decision Theory and Multi-Agent Planning, Springer Wien/New York.

Report: Rank S., Petta P., Trappl R. (2004): Features of Emotional Planning in Software Agents. Technical Report, Österreichisches Forschungsinstitut für Artificial Intelligence, Wien, TR-2004–23.

Features of Emotional Planning in Software Agents

Extended Abstract

Stefan Rank[‡], Paolo Petta[†‡], Robert Trappl[†‡]
[†] Dept. of Medical Cybernetics and Artificial Intelligence,
Medical University of Vienna
[‡] Austrian Research Institute for Artificial Intelligence*,
Vienna, Austria
{stefan.rank@ofai.at, paolo.petta@meduniwien.ac.at, Robert.Trappl@meduniwien.ac.at}

1 Introduction

It can be argued that emotions are an essential element of intelligence; they are certainly relevant for cognition and action in humans. We believe that software agents can benefit from explicit consideration of emotional processes in the design of their architecture. Our work extends the foundation laid in work on the TABASCO framework [Petta 1999, Petta 2003] by addressing the relation of planning capabilities and emotional processes in agents that are resource-bounded and situated in complex (rich, social, dynamic, and partially observable) environments: planning is considered as a separate module of the agent, that can — but need not — be consulted during execution. We first introduce our approach towards integration of planning in emotional situated agents, analysing the interfaces of continuous planners and the emotion process as conceptualised by cognitive appraisal theories. Next, we cover some implemented systems that already integrate aspects of emotion theories and planning. We conclude with a summary of our findings on the interrelation of planning and emotion processes.

2 What kind of planning?

Decision Theoretic Planning [Blythe 1999, Boutilier et al. 1999] tries to overcome assumptions about the environment shared by classical planners, including binary ("all-or-nothing") goal achievement, and infallible actions. Decision Theory explains rational choice on the grounds of *uncertain knowledge* and *fallible action.* The basic building blocks of a decision are a *preference function* over action outcomes and a *probability distribution* of the *expected possible outcomes* of execution of a given action. In combination, they provide a measure of *expected utility*, leaving the deciding agent with the "simple" task of determining the *maximum expected utility* of its current options. Complex environments pose several additional planning challenges: knowledge about the environment might be incomplete or wrong; actions may fail; results of actions may

*The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry for Education, Science and Culture and by the Austrian Federal Ministry for Transport, Innovation and Technology. This research is carried out within the European Network of Excellence Humaine (IST-2002-2.3.1.6 507422) with support from the Austrian Funds for Research and Technology Promotion for Industry (FFF 808818/2970 KA/SA).

come undone; goals may be achieved to different (partial) degrees; and there are multiple conflicting goals to be pursued. As a consequence, goals can be of several different types, such as avoiding, achieving, or maintaining a state of the world, as well as verifying it. In order to cope with the dynamics of such an environment — including exogeneous effects —, planner and executive have to run concurrently, or at least in some interleaved fashion.

In the design of such situated agents, we consider *motivation-driven continuous planning systems* [Avradinis & Aylett 2003], and especially *hybrid systems*: behaviour-based architectures [Brooks 1991] that employ a planner as a separate supportive resource or layer that is not always needed for the agents' functioning (cf. [Ferguson 1992, Jung & Fischer 1998]). In this scenario, plans are entities that are mainly *used*, rather than *constructed*. This view of plans is informed by the notion of *plans-as-commitments* and *plans-as-communication* — structures that rely on a substantial amount of run-time contextual interpretation for execution. The plans-as-commitments view regards established *intentions* as filters for new options to be considered in a continuous execution environment [Pollack 1992, Pollack & Horty 1999], thereby containing the need for reasoning. Commitments also influence the estimation of possible costs of new options [Horty & Pollack 2000], e.g., when a new option requires only a minor modification of a current plan. The plans-as-communication view [Agre & Chapman 1991] introduces abstract plans that do not immediatly control execution. Instead, *interpretations* bind plans to substantial current contextual information, achieving effective refinement and re-use. Such indirect employment of (meta-)control information is compatible with hybrid architectures that use deliberative modules as *support* for reactive ones: an interpretation might e.g. cause the (re-)configuration of an architectural layer.

An architecture for situated agents has to coordinate influences arising from the environment and those stemming from the agent itself. This is another reason why a planning module for such an architecture has to run concurrently — either continuously, or with support for being suspended. To be safely stopped and (re-)started at arbitrary times, a planning module should support partial plans, be capable of plan adaptation[1], and have anytime characteristics. In characterising the interface of such a planning module, we distinguish the parameters from management facilities. The latter comprise all commands to steer its activity, e.g., start, stop, suspend, and inspect. Beyond goals and beliefs, inputs need to include utilities and costs of activities — both of which possibly can in turn be made up of smaller building blocks —, as well as general and domain-specific heuristics, and criteria of success. Outputs do not only consist of plan structures of different types, but also include the options abandoned because determined to be unusable at the current moment. In particular, such abandoned plans (and their unfulfilled preconditions) can be instrumental in monitoring the environment for new options. Further, metering and profiling information includes the options and number of alternatives considered, the level of detail of the plans returned, and the expended effort.

3 What kind of emotion?

Out of the numerous existing theories, the *cognitive appraisal theories of emotion* have experienced a substantial consolidation and empirical validation in the recent past. In this model, emotional experiences result from evaluations of the subjective significance of construed situations and events, according to specific dimensions or criteria [Frijda 1986, Roseman et al. 1996, Roseman & Smith 2001, Ellsworth & Scherer 2003]. Emotion is delineated as a *process* (rather than the descriptive characterisations of emotions in dimensional or categorical models) we adopt as crucial element for the survival of agents situated in complex social environments.

[1] The classic term "plan repair" reminds of the original static and closed environments, where modification of plan structures was the exception, rather than the rule.

According to [Ellsworth & Scherer 2003], major subjective dimensions of stimulus appraisal include: *novelty* and *pleasantness* (gatekeepers of the emotion system, deciding whether a stimulus is noteworthy); the *conduciveness* of a stimulus to the agent's *concerns* (i.e., motivational constructs, subsuming needs and goals); the *potential to cope* with the situation; and the *social implications* (an event's relation to social norms and values upheld by the agent). The emotion process instigates a pervasive synchronised effort of the agent's resources, including the planning subsystem, to adapt to and cope with the perceived subjective significance of the eliciting event: to attend to or avoid, support or obstruct, pay attention to or ignore, etc., the *perceived likely meaning (implications)* of the stimulus event. This *action tendency change*, the essence of an emotional episode, is subjectively experienced, and overtly disclosed in expressive behaviour. The term *coping* denotes the ensuing activities to map this actual motivating impulse to subsequent action: action tendencies may be inhibited or reinforced; planning parameters updated; current plans-as-commitments revised; or new goals established; so as to adapt to the change subjectively assessed.

Accordingly, main outputs of emotions are *action control* precedence signals (e.g., interrupts) that try to influence different parts of the agent, as well as the emotion process itself. Further outputs include *internal awareness* and *overt manifestations* (e.g., facial expressions; gestures; exclamations: the very *expressiveness* of behaviour) of the action tendency change proposed. The interface of the emotion process thus provides a structured subjective interpretation of situational changes along the appraisal criteria; preliminaries for adaptation in terms of suggested action tendency changes (including action control influences); subjective awareness; and overt signalling under limited voluntary control. *Situational parameters* of the emotion process consist of the agent's motivational constructs; expectations; and sensed actual situational changes. *Dispositional factors* — side-conditions of the emotion process — include: the coding categories to appraise aspects of situational change and the readiness to use them; response thresholds and response modes (e.g., favouring antagonistic or protective behaviour); and previous experiences with different types of events. Finally, *regulatory* influences — such as *display rules* — can be seen to implement management functionalities by constraining and shaping: the appraisal process; action tendency changes and admissible ways to cope with and implement them; and allowable ways of expression.

This short analysis already points to a number of noteworthy aspects of the interplay of planning capabilities and emotional functionalities. We next take a brief look at some related work.

4 Emotional planning architectures

The APOC framework [Scheutz 2001] was used to compare and contrast the evolutionary values of affective and deliberative control components as independent additions to a purely reactive agent design. Their performance in a foraging scenario was compared, with affect being represented by explicitly reified state variables (e.g., for hunger or thirst), and deliberative control comprising memory and route planning modules. In some environments, the simulated affective species proved more likely to survive than the deliberative one, and would also be more performant than variants combing merely competing affective and deliberative capabilities. Besides emphasising the inevitable and essential impact of environmental factors on actual performance, this works thus also indicates the importance of architectural integration of deliberative and affective functionalities. Matthias Scheutz' work on APOC was greatly influenced by the CogAff project led by Aaron Sloman. His pursuit of a design-based approach for cognitive agent architectures includes also an effort to anchor the definitions of emotional concepts in an architecture-based theory of agent designs [Sloman 2004], where terminology refers unequivocally to components and their interconnections, functions, and forms of representation.

A very advanced model combining planning and affect implemented in running systems is the emotional framework EMA [Gratch & Marsella 2004]. EMA is aimed at a domain-independent way of modelling the mechanisms underlying human emotion. It relies substantially on concepts of planning for the domain-independent causal interpretation of the world. Here, the term coping is used for all activities that influence the system due to appraisals. Coping is viewed as the inverse of appraisal: the identification and influencing of the believed causes for appraisal. EMA covers a broad variety of coping strategies (e.g., "mental disengagement") that have been mapped to decision-theoretic planning techniques (e.g., by lowering the utility of the responsible goal). Finally, the starting point of our own line of work are the integration of insights from psychological theories on emotion with results from the areas of cognitive robotics and intelligent agents, for a principled generation and understanding of expressive behaviour. Example implementations of TABASCO, our framework for tractable appraisal-based architectures for situated cognisers [Staller & Petta 1998, Petta 2003], include the invisible person, an interactive exhibit [Petta 1999], and ActAffAct[Rank 2004], a dramatic environment test-bed.

5 Towards integrating planning and emotion

Our aim is to identify and implement a set of functional components for a system to display "believable" emotional action [Bates 1994] in a virtual environment. In the present paper, we focussed our analysis on the contributions that planning capabilities can provide for and receive from emotional processes and vice versa, against the particular background posed by agents situated in complex settings. Findings from the evolution of planning technologies on the one hand and state-of-the-art theory in cognitive emotion psychology on the other appear to show substantial compatibilities (if not direct matchings), as evidenced by the comparison of the identified notions of goals vs. concerns; beliefs and world facts vs. components of subjective meaning structures; utilities vs. hedonic values and measures of coping potential; costs vs. impulse strengths; and heuristics vs. response modes. Computational theories have long been providing important contributions to theory building in cognitive psychology. We believe that insights from emotion research, such as those that led to concepts such as *concerns* and *action tendencies*, can now in turn lead to fruitful innovation in the computational community. Within its assistance in helping to come to terms with the full implications of *situatedness*, *boundedness*, and *sociality*, we see a particular potential of this line of research in contributing to overcome the long-standing historic impasse of an overly big reliance on economic models (cf. the substantial evidence of the disagreement of human behaviour and decision theory).

We thus assert that situated agents can profit from, both, the more abstracted and objectified — i.e., *sharable* — perspective of planning, *and* the subjective and grounded *current* evaluations in the emotion process; and that elements from both views are in fact *required* to achieve a whole architectural design. Successful synthesis of the two perspectives, however, necessitates of a deeper consolidation and integration of their functionalities, with reconceptualisations beyond what has been realised to date (cf. the evolution from ACRES [Swagerman 1987] to *Will* [Moffat 1997]). For this reason, the possibilities disclosed by moving from a continuous planner exerting centralised control towards planning as a resource in a concurrent hybrid interaction architecture are of special interest to us. Our work continues with a further clarification of the relations of the basic concepts from the two domains identified so far with further elements of the architecture and their grounding in the agent's physical and social environment, in collaboration with other emotion researchers within the European 6^{th}FP Network of Excellence, Humaine[2]. For empirical testing, the results are implemented in revisions of control architectures developed previously [Petta 2003, Rank 2004].

[2] http://emotion-research.net

References

[Agre & Chapman 1991] Agre P.E., Chapman D.: *What are plans for?*, Technical Report: AIM-1050a, Massachusetts Institute of Technology, Cambridge, MA, USA, 1991.

[Avradinis & Aylett 2003] Avradinis N., Aylett R.: Agents with no aims: Motivation-driven continuous planning, Rist T. et al. (eds.), *Intelligent Virtual Agents*, LNCS 2792, Springer Berlin/Heidelberg, 2003.

[Bates 1994] Bates J.: The Role of Emotion in Believable Agents, *Communications of the ACM*, Special Issue: Intelligent Agents, **37**(7), 122–125, 1994.

[Blythe 1999] Blythe J.: Decision-theoretic Planning, *AI Magazine*, **20**(2):37–54, 1999.

[Boutilier et al. 1999] Boutilier C., Dean T., Hanks, S.: Decision-Theoretic Planning: Structural Assumptions and Computational Leverage, *Journal of Artificial Intelligence Research*, **11**(1):1–94, 1999

[Brooks 1991] Brooks R.A.: Intelligence without Representation, *Artificial Intelligence*, Special Volume: Foundations of Artificial Intelligence, **47**(1-3):139–159, 1991.

[Ellsworth & Scherer 2003] Ellsworth P.C., Scherer K.R.: Appraisal Processes in Emotion, Davidson R.J. et al. (eds.), *Handbook of Affective Sciences*, Oxford University Press, Oxford New York, 572–595, 2003.

[Ferguson 1992] Ferguson I.A.: *TouringMachines: An Architecture for Dynamic, Rational, Mobile Agents*, University of Cambridge, UK, Ph.D. Thesis, 1992.

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

[Gratch & Marsella 2004] Gratch J., Marsella S.: A Domain-independent Framework for Modeling Emotion, *Cognitive Systems Research*, **5**(4):269–306, 2004.

[Horty & Pollack 2000] Horty J.F., Pollack M.E.: *Evaluating New Options in the Context of Existing Plans*, Institute for Advanced Computer Studies, University of Maryland, College Park, MD, 2000.

[Jung & Fischer 1998] Jung C.G., Fischer K.: *Methodological Comparison of Agent Models*, DFKI Research Report-98-01, DFKI, Saarbrücken, Germany, 1998.

[Lazarus 1991] Lazarus R.S.: *Emotion and Adaptation*, Oxford Univ. Press, London/Oxford/New York, 1991.

[Moffat 1997] Moffat D.: Personality Parameters and Programs, in Trappl R., Petta P. (eds.), *Creating Personalities for Synthetic Actors*, LNAI 1195, Springer-Verlag, Berlin/Heidelberg, 120–165, 1997.

[Myers 1999] Myers K.L.: CPEF: A Continuous Planning and Execution Framework, *AI Magazine*, **20**(4), 63–70, 1999.

[Ortony 2003] Ortony A.: On Making Believable Emotional Agents Believable, in Trappl R. et al. (eds.), *Emotions in Humans and Artifacts*, MIT Press Cambridge, MA/London, UK, 189–212, 2003.

[Petta 1999] Petta P.: Principled Generation of Expressive Behavior in an Interactive Exhibit, in Velásquez J.D. (ed.), Agents'99 Workshop: *Emotion-Based Agent Architectures (EBAA'99)*, Seattle, WA, USA, 94–98, 1999.

[Petta 2003] Petta P.: The Role of Emotions in a Tractable Architecture for Situated Cognizers, in Trappl R. et al. (eds.), *Emotions in Humans and Artifacts*, MIT Press Cambridge, MA, London, UK, 251–288, 2003.

[Pollack 1992] Pollack M.E.: The Uses of Plans, *Artificial Intelligence*, **57**(1), 43–68, 1992.

[Pollack & Horty 1999] Pollack M.E., Horty J.F.: There's More to Life Than Making Plans: Plan Management in Dynamic, Multiagent Environments, *AI Magazine*, **20**(4), 71–84, 1999.

[Rank 2004] Rank S.: *Affective Acting: An Appraisal-based Architecture for Agents as Actors*, Vienna University of Technology, M.S. Thesis, 2004.

[Roseman et al. 1996] Roseman I.J., Antoniou A.A., Jose P.E.: Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory, *Cognition and Emotion*, **10**(3):241–277, 1996.

[Roseman & Smith 2001] Roseman I.J., Smith C.A.: Appraisal Theory: Overview, Assumptions, Varieties, Controversies, in Scherer K.R. et al. (eds.), *Appraisal Processes in Emotion: Theory, Methods, Research*, Oxford University Press, Oxford New York, 3–19, 2001.

[Scheutz 2001] Scheutz M.: Affective vs. Deliberative Agent Control, *Proceedings of the AISB 2001 Convention: Emotion, Cognition and Affective Computing*, March 21–24, 2001, University of York, UK, 1–10, 2001.

[Sloman 2004] Sloman A.: What Are Emotion Theories About?, Hudlicka E., Cañamero L. (eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations*, Papers from 2004 AAAI Spring Symposium, Technical Report SS-04-02, AAAI Press/MIT Press, Cambridge/Menlo Park, 128–134, 2004.

[Staller & Petta 1998] Staller A., Petta P.: Towards a Tractable Appraisal-Based Architecture for Situated Cognizers, in Cañamero D. et al. (eds.), *Grounding Emotions in Adaptive Systems*, Workshop Notes, 5^{th} Intl. Conf. of the Soc. for Adaptive Behaviour (SAB98), Zurich, Switzerland, 56–61, 1998.

[Swagerman 1987] Swagerman J.: *The ARtificial Concern Realization System ACRES: A Computer Model of Emotion*, University of Amsterdam, The Netherlands, Ph.D. Thesis, 1987.