

Humaine Workshop
“Theories and Models of Emotion”, June 17-19, 2004

“Identifying theoretical problems”
Morning Session, June 18, 2004

Synthesis

Paolo Petta (Session Chair)
Austrian Research Institute for Artificial Intelligence, Vienna, Austria
(paolo@oefai.at)

Introduction

The purpose of this session was not so much the presentation of the different systems themselves, but rather to use these cases as examples to illustrate how theory and application-oriented efforts relate in practice, and what issues are faced when looking for a sound theoretical fundament for application design and when trying to operationalise and implement identified theories. These statements grounded in own research experience aimed at helping set off the discussion on how to improve the connections between the research community in psychology and those concerned with the design and development of emotion-oriented systems. Consequently, the contributions focussed on discussing both, matches and mismatches, between theoretical foundations and practical necessities – before; during; and after completing work on the respective systems – including the pointing out of topics for which no (satisfactory) theoretical support could be identified. To best exploit the precious limited time at the workshop, the presenters made available descriptions of the coverage of their presentations along with background literature on the humaine portal in advance, and the session chair invited all members and friends of humaine registered for the workshop to consult this material to acquire some working knowledge about the systems to be presented, and prepare questions, comments and suggestions for topics of discussion.

A brief survey of the presentations

Along the line of argument exposed in the introduction, the following is not meant to reproduce the full content of the different presentations, but to relate some important issues and topics that emerged during the session to the presentations where these were touched upon.

The morning session began with a joint presentation by Ana Paiva and João Dias from INESC-ID and IST in Lisbon, Portugal. In her introductory remarks, Ana Paiva brought up the basic system designer’s issue of having to identify which psychological theory fits best for which task. She illustrated mismatches between the scope of emotion theories and the actual requirements and needs of applications using the example of the OCC model, referring not only to limitations of the emotion model, but also discussing project scenarios in which it eventually turned out that little – if any – of this kind of emotion theory was needed at all. With reference to most recent work in the context of the VICTEC IST project, Ana Paiva hinted at the failure of an attempt to encode Klaus Scherer's component model of appraisal theory, due to of a lack of control over the dynamics of the implementation that ended up getting locked in small behaviour cycles of indefinite duration. At the same time, it remained unclear whether these problems were due to limitations or weaknesses in the theory, or rather came about because of insufficient coverage of implemented aspects – e.g., the relevance and

downright necessity of pervasive dual control, of excitatory and regulatory elements at different levels would come to mind (cf. e.g., Scherer 1982, Frijda 1986, Moffat et al. 1993, Gross 1998, Gross 1999, Frankel 1999, Hudlicka 2001, Leary 2003). The presentation of the VICTEC system also illustrated the practical relevance of parametric models, where behaviour parameterisation allows to use a single model (control architecture) to generate a range of different behaviour types that are perceived as different personalities.

The joint presentation by Brigitte Krenn (OFAI -Vienna, Austria) and Marc Schröder (DFKI – Saarbrücken, Germany) used the recently completed IST project NECA to emphasise the significance of achieving expressivity in embodied conversational agents (ECAs). Here, Brigitte Krenn contrasted the system designer’s need for characterisations of dynamics with Paul Ekman’s emotion categories, which provide only a static model of facial expressions. There furthermore is a strong need for information on the impact of facial expression changes and dynamics on the perceived emotion, as well as a lack of information regarding the roles of different modalities in multi-modal presentation. Further topics raised in the presentation included the confrontation of exaggerated presentation against natural system behaviour, or of staged communicative performance vs. situated task/goal-directed behaviour. Focussing further on today’s engineering challenges, the presentation covered the issue of the subjectivity of annotations of multi-modal content, along different dimensions that include the choice of taxonomies as well as inconsistent usage patterns of shared vocabulary terms. Another major topic brought forward in this talk regarded the mapping between different theories: to which extent are they equivalent in terms of covering the same ground, where do complementarities exist? What about fundamental compatibilities of different theories, where results may be based not only on different conceptualisations for different purposes (e.g., Solomon 2001, Pfeifer 1994, Wehrle & Scherer 1995, Davidson et al. 2003, Segalowitz & Cohen 2003, Parkinson 1996, Keltner & Gross 1999, Levenson 1999, Salovey 2003), but also be derived from and pertain to substantially different species (cf. e.g. Feshbach 1987, Plutchik 1987, Davis 1994, Preston 2002; vs. the recent discussion about important commonalities and differences between emotion-related brain structures in rodents and humans, e.g. Holland & Gallagher 2004 vs. Irwin et al. 2004)?

At the beginning of his presentation, Björn Hartmann (IUT Montreuil – Univ.Paris VIII, Paris, France) characterised the wide span faced in the implementation of emotion-based systems employing ECAs, ranging from the control of geometrical angles and polygons to the levels of theories, which in turn may provide more abstract surface or shallow models on the one hand, or deeper models on the other (cf. the references given on the “State of the Art” slide of the presentation). The focus of the presentation lied with the illustration of the significant progress made towards enabling emotional facial displays in ECAs through expressivity control, overcoming the limitations of alternatives based on comparatively simple statistical models (e.g. Perlin 1997). New avenues on the horizon include shifting away further still from animation approaches inherited from the computer graphics area - such as inverse kinematics – and approaching the kind of perceptual control exerted in living beings. This entails providing increasingly more detailed models that include physical simulations that shape/constrain the high-dimensional space of degrees of freedom (cf. e.g. Pfeifer 2004) Björn Hartmann succeeded in conveying a very vivid and clear picture of the tension between the applications asking for clear-cut, slim and deterministic models for implementation of control, and the broadness of the natural phenomenon. In an intervention in this context, Klaus Scherer also referred to the discussion surrounding the theoretical notions of personality and dispositions. Both of these latter issues were also topicalised in the subsequent talks.

In the presentation of results and insights gathered in the European IST project ERMIS, Ellen Douglas-Cowie (Queen’s University, Belfast, UK) addressed the tasks of elicitation of emotion and emotion coloured discourse. There, she pointed out how people have heightened and simplified cartoonish images of it, which may even be an inbuilt property of humans. With respect to the generation of behaviour for the purposes of elicitation of emotion, the effectiveness of the “ELIZA-style architectures” (Weizenbaum 1966) employed in the SALAS (“Sensitive Artificial Listeners’ Association”) System in eliciting emotional signals from users was highlighted. SALAS is capable of portraying different personalities and control paradigms (compelling, permitting,...). A number of significant issues related to the recording of emotional signals were raised, many of which are scheduled to be addressed in detail in the forthcoming first humaine summer school. Also pointing ahead – to the WP4 workshop – tasks and related theoretical issues in the description of emotional content of acquired material were discussed.

The presentation of the synthesis part of ERMIS pointed at further areas of extant application needs: theoretical knowledge is lacking to both, model the output, and to guide and inform the generalisation and mapping mechanisms that transform and merge the content descriptions resulting from the analysis procedures (e.g., of facial expressions). In the ERMIS project, neurofuzzy techniques are used as “glue technology” to connect the different levels.

The importance of theoretical guidance for the analysis of scenarios and the development of descriptors was underscored in the joint presentation by Lola D. Cañamero, Orlando Avila-Garcia, and René te Boekhorst (Univ.Hertfordshire, Hatfield, UK). Experiments with emotional behaviour control of mobile robots via hormonal modulation of cognition and action were used to illustrate the modelling of homeostatic motivations and the analysis of designs for distributed motivation arbitration, in particular different trade-offs of winner-take all and voting-based (majority) schemes. With reference to Braitenberg’s “Law of Uphill Analysis and Downhill Invention” (Braitenberg 1984, p.20), difficulties in identifying structural conceptual components of situated dynamic behaviour as engendered even by comparatively simple control architectures were discussed. René te Boekhorst presented solution approaches from ethology to tackle these issues.

In the last presentation, Christine Lisetti (Eurécom, Sophia Antipolis, France – invited contribution from outside humaine) presented her MAUI (a Multimodal Affective User Interface) system framework. As in ERMIS, this system is aimed at recognising a user’s emotions by sensing various user-centred modalities (here: visual, kinaesthetic and auditory). The comprehensive framework illustrating the broad range of component technologies and modelling approaches involved raised similar questions – e.g. with respect to the emotion taxonomy used by the system and the algorithms used to perform the mapping from sensed quantities to this taxonomy – as previous presentations.

Analysis

Central to humaine’s Workpackage 3, “Theories and Models” is to assess and improve over the current status of the relation between what (psychological) emotion theories have to offer on the one hand and the needs and aims of applications on the other. To this effect, the first morning session of the workshop succeeded in highlighting a number of important and typical issues that characterise the current state of the area of affective computing. In the subsequent discussions and working sessions, some relevant sources of information could be identified and first decisions on how to deal with these challenges were agreed upon.

Overall, important issues are indicated by the various “failures” in attempts to implement and use available theoretical models in a direct and straightforward fashion with an engineering mindset. Possible causes of such failures include:

- a basic misunderstanding of the scope and status of the psychological theories by the application designer, leading to attempts to utilise theories in contexts where they are not appropriate;
- partial implementations, that omit constituents of the theory that may have been deemed irrelevant for the application or that are only referred to or omitted altogether in the literature consulted by the development team;
- partial theories, that omit constituents and level of detail necessary for implementation;
- elements missed by implementations that are presupposed by theory (often tacitly) or difficult or impossible to provide because of various limitations (such as lack of prescriptive models of sufficient detail; issues related to human interface technologies required; lack of content data such as knowledge and case bases of sufficient mass);

A special case that appears to deserve being singled out is the popularity of the OCC model (Ortony et al. 1988, Ortony 2003) that seems to be misunderstood at times to be of just about universal applicability. Issues again arise from the scope of theory, often forcefully applied to purposes as not first intended by the authors (this applies in particular to Ortony et al. 1988; see Ortony 2003 for some extensions towards a first-person, generative model that covers more of the elements of the main appraisal theories); the actual needs of the application at hand then often end up being left with no (or in any case, dissatisfactory) coverage. This theory furthermore presupposes significant infrastructure (that includes e.g. the issue of modelling of standards and norms, as well as related evaluation processes – introducing dependencies on ongoing debates such as on the status of the deontic reasoning vs. deontic logic; see e.g. Gratch & Marsella 2004 for a recent critique).

More generally, a central design issue was addressed in practically all contributions to the session: the importance of identifying theoretical contributions of a level of sophistication (or, rather, simplicity) that matches the application at hand. Theoretical overdesign has to be avoided, which manifests itself by forcing designers to simplify and abstract elements in ad-hoc ways in implemented architectures (with the extreme example of references to “consciousness”). The examples covered in the talks sampled the design space at different degrees of abstraction, from simulated hormones for modulation of robotic control, over the generative technologies employed in Greta (C. Pelachaud et al.) and NECA (B. Krenn et al.), to the interactive social VicTec system (A. Paiva et al.).

Another issue that has to be considered carefully is whether a given applications aims mainly at reactive; natural; impulsive;... or exaggerated; staged; controlled;... behaviour: Variants on this topic also include the necessary distinction of immediate generation of emotion from further elaboration on emotions (as e.g. indicated in the “policy box” management elements of the architectural diagram sketched in Ball 2003). This again immediately begs the question whether theories of emotions can be expected reasonably to be the prime contributors of significant background for the design and engineering efforts, or whether it would not be more advisable to look elsewhere first, e.g. at skills and recipes of communication and persuasion. It thus may be preferable at times to rely on shallow models from empirical studies (that also accommodate strong influences of elements such as cultural conventions). First most helpful indications to this effect were provided at the workshop by Klaus Scherer with his references to “buried” existing literature to be brought to the attention of humaine.

The issue of obtaining and tuning parameters is another strong requirement of applications that ultimately cannot get by with merely qualitative descriptors. Here, even when a model of emotional processes is really at the core of an application, it has to be clarified whether and to which extent an expectation would be justified that once an architecture has been instantiated by following closely psychological theory, all it then takes is “just a matter of calibrating parameters”. Caution has to be taken not to conflate different and unrelated topics, expecting emotion-related parameters to provide e.g. dynamical stability that in the natural role model results from the interplay with numerous other subsystems. Artificial systems that are to provide functionalities of high complexity while remaining controllable in detail necessarily will have to delineate clearly the scope of the (grounded) theory-abiding parts and those following engineering needs and principles, as well as to define their interplay.

On the other hand, issues such as the need to overstate inherent in communication could also be seen as a non-problem (or at least a derived problem pertaining elsewhere) that arises as a consequence of simplified modelling of two-way interaction (that would always include negotiation and feedback as constituent part) as sequenced turns of one-way emission/sending. Seen this way, the real issue is the need to supersede the notions of control and action with approaches placing a higher emphasis on distributed coordination and interaction.

The way ahead

It is crucial to acknowledge that the important issues in the working area of humaine, such as the ones mentioned in this short document, do exist not only because of current shortcomings in the individual areas, such as engineering, computer science, and psychology: Contributions to solutions cannot e.g. be expected to be provided from the psychological end alone, for such activities are met with strong scepticism in that scientific community. In the discussion, Stacy Marsella¹ contributed most helpfully the vivid metaphor equating such endeavour to “suicide of a psychologist”. This tension between limitations of what psychological theory can be expected to contribute and what application engineering does need is characteristic of a current void which humaine is arranged to help bridge, with interdisciplinary collaborations covering different perspectives and at different levels of analysis (in a very rough ordering, examples from the workshop session would include statistical ethological behaviour analysis; neurofuzzy models relating sensor-provided data to elements of abstract conceptualisations such as (EM)FACS coding; and the dyad, group, and social levels).

In this effort, it is of decisive importance to assess critically the viability of combinations of models and theories from different backgrounds, so as to avoid haphazard plumbing - e.g., to recognise the status of BDI models of agent control as derived from one particular model of practical reasoning (as opposed to more general and encompassing cognitive theories such as ACT-R or SOAR), or the differences between the notions of plans and planning in computer science and psychology (and sociology). Further methods of work include reconciliation (mapping) of and across theories, and the synthesis of new, coherent models that integrate contributions from different sources into designs suited to engineering purposes. The range of affective competence asked for in application settings as presented in this session require further efforts on componential generative models of emotions, so as to progress beyond the limited and often fixed range of (e.g. “archetypal”) emotions of the vast majority of today’s systems. At the same time, progress/development in existing frameworks should be tracked (as with the evolution of the OCC model).

Of similar and straightforward relevance is the assessment and acknowledgment of the possibilities and limitations of the state of the art in engineering - with respect to both, analysis: e.g. the recognizing; tracking; and segmentation of faces under largely unconstrained

¹ Arguably, himself indictable of such sacrificing behaviour, for the good of the nascent community of affective scientists.

conditions; the interpretation of facial expressions; and synthesis: e.g. of facial expressions, voice, gaze, and gesturing in ECAs. The presentations of this first workshop session addressed these topics, when referring to the different examples and approaches taken therein. Component technologies were covered in the context of projects such as ERMIS and Greta, that follow a bottom-up approach; engineering issues related to the top-down design of complete applications and whole architectures were discussed with respect to different applications that were developed by INESC-ID/IST; application scenarios for the MAUI framework; or the design of the robotic control architectures at the University of Hertfordshire.

To make progress, it will be important to work towards consolidating available evidence regarding the role(s) of emotions: the presentations in the session proposed and contrasted the variants of modulation (L.Cañamero); control (A.Paiva, B.Krenn); and a more diffuse (and perhaps less simplifying) being “all over the place” in analysis (R.Cowie, S.Kollias). Within this broad range, the nature (identity, compatibility) of the concepts of “emotion” needs to be assessed and documented. With respect to the notion of control in affective applications, it appears to be important to assess critically its pertinence and relevance in applications of this type: as already referred to above, the impact of changing the default approach from control to regulation, from sending/acting to diachronic communicating/*interacting* (inherently bi-directional, embedded in context, with history and expectations), from linear and monolithic architectures to designs that reflect asynchronicity and distributedness, needs to be followed through. Also, there is the need to reflect on insights available from situated agent research, as exemplified by the contrast of the statement “*oddly enough, an ELIZA-kind of approach is enough*” in the presentation of the ERMIS project with the results of the analysis of the computational power of situated automata (Agre & Chapman 1987; Rosenschein & Kaelbling 1996; Wegner 1997).

Another perspective on the issue of the relation of emotion and interaction (perception, situatedness) was provided in the way emotional appraisal has been (ab)used in implemented systems, e.g. as planning operator or immediate goal selector – in contrast to the conceptualisation of emotional appraisal as subjective evaluation in psychological theory. At the same time, the presentations demonstrated the range of important insights available from the different backgrounds in the engineering and technical sciences. For example, the body of work of computational linguists on dialogue systems has led to e.g., a deeper understanding of the impact of structured longer-term context; to theoretical progress from the notion of speech acts to the notion of dialogue acts as unit of analysis; and, in the context of ECAs, to an appreciation of the roles, dependencies, and interplays of modalities and multi-modalities.

Conclusion

To summarise, the success of the (short) session is not only documented in the substantial material that could be collected (if still only in part) in this first synthesis, whose coverage is strictly limited to this first working session of the workshop. The engaged ways the theoretical issues first raised here were carried over into and expanded upon in all the subsequent sessions of the workshop – along with their grounding in showcase application settings and the related webs of argumentation – are just as important evidence for how the event managed to meet the expectations, for the purposes of the workplan of WP 3 in particular, and humaine overall. As chair of the session, I would like to extend my thanks to all who contributed to the presentations, as well as to all participants, who contributed in turn and then readily picked up and perpetuated the momentum generated here.

References:

- Agre P.E., Chapman D.: Pengi: An Implementation of a Theory of Activity, in *Proceedings of the 6th National Conference on Artificial Intelligence*, Morgan Kaufmann, Los Altos/Palo Alto/San Francisco, pp.268-272, 1987.
- Ball E.: A Bayesian Heart: Computer Recognition and Simulation of Emotion, in Trapp R., et al.(eds.), *Emotions in Humans and Artifacts*, MIT Press Cambridge, Massachusetts London, England, pp.303-332, 2003.
- Braitenberg V.: *Vehicles. Experiments in Synthetic Psychology*, MIT Press, Cambridge/Boston/London, 1984.
- Davidson R.J., Scherer K.R., Goldsmith H.H.: Introduction: Neuroscience, in Davidson R.J. et al. (eds.), *Handbook of Affective Sciences*, Oxford University Press, Oxford New York, pp.3-7 (Chapter 1), 2003.
- Davis M.H.: *Empathy: a social psychological approach*, McGraw-Hill College, 1994.
- Feshbach N.: Parental empathy and child adjustment/maladjustment, in: Eisenberg N., Stryker J. (eds.): *Empathy and its Development*, Cambridge University Press, Cambridge, pp. 271-291, 1987.
- Frankel C.B.: Such order from confusion sprung: adaptive competence and affect regulation, Pacific Graduate School of Psychology, Palo Alto, CA, USA, Ph.D. Thesis, 1999.
- Gratch J., Marsella S.: A Domain-independent Framework for Modeling Emotion, *Cognitive Systems Research*, in press, 2004.
- Gross J.J.: The Emerging Field of Emotion Regulation: An Integrative Review, *Review of General Psychology*, 2(3):271-299, 1998.
- Gross J.J.: Emotion Regulation: Past, Present, Future, *Cognition and Emotion*, 13(5):551-573, 1999.
- Holland P.C., Gallagher M.: Amygdala – frontal interactions and reward expectancy, *Current Opinion in Neurobiology* 14(2):148-155, 2004.
- Hudlicka E.: Modeling Affect Regulation and Induction, in Canamero L.D., *Emotional and Intelligent II: The Tangled Knot of Social Cognition - Papers from the 2001 Fall Symposium, North Falmouth, MA*, AAAI Press, Menlo Park, CA, Technical Report FS-01-02, pp.51-56, 2001.
- Irwin W., Anderle M.J., Abercrombie H.C., Schaefer S.M., Kalin N.H., Davidson R.J.: Amygdalar interhemispheric functional connectivity differs between the non-depressed and depressed human brain, *NeuroImage*, 21(2):674-686, 2004.
- Keltner D., Gross J.J.: Functional Accounts of Emotions, *Cognition and Emotion*, 13(5):467-480, 1999.
- Leary M.R.: The Self and Emotion: The Role of Self-Reflection in the Generation and Regulation of Affective Experience, in Davidson R.J., et al.(eds.), *Handbook of Affective Sciences*, Oxford University Press, Oxford New York, pp.773-786 (Chapter 40), 2003.
- Levenson R.W.: The Intrapersonal Functions of Emotion, *Cognition and Emotion*, 13(5):481-504, 1999.
- Moffat D., Frijda N., Phaf H.: Analysis of a model of emotions, in Sloman A., Hogg D., Humphreys G., Ramsay A. (eds.): *Prospects for artificial intelligence*, IOS Press, Amsterdam, pp.219-228, 1993.
- Ortony A.: On Making Believable Emotional Agents Believable, in Trapp R., et al.(eds.), *Emotions in Humans and Artifacts*, MIT Press Cambridge, Massachusetts London, England, pp.189-212, 2003.
- Ortony A., Clore G.L., Collins A.: *The Cognitive Structure of Emotions*, Cambridge University Press, Cambridge, UK, 1988.
- Parkinson B.: Emotions are social, *British Journal of Psychology*, 87:663-683, 1996.

- Perlin K.: Layered Composition of Facial Expression, *ACM SIGGRAPH 97 Visual Proceedings: The art and interdisciplinary programs of SIGGRAPH'97, Los Angeles, CA, USA*, 226-227, 1997.
- Pfeifer R.: The Fungus Eater approach to emotion: a view from artificial intelligence, *Cognitive Studies*, 1:42-57, 1994.
- Pfeifer R., Iida F.: Embodied Artificial Intelligence: Trends and Challenges, in Iida F., et al. (eds.), *Embodied Artificial Intelligence*, Springer-Verlag Berlin Heidelberg New York, Lecture Notes in Artificial Intelligence 3139, pp.1-26, 2004.
- Plutchik R.: Evolutionary bases of empathy, in: Eisenberg N., Staryer J. (eds.): *Empathy and its Development*, Cambridge University Press, Cambridge, pp. 38-46, 1987.
- Preston S.D., Waal F.B.M.de: Empathy: Its ultimate and proximate bases, *Behavioral and Brain Sciences*, 25:1-72, 2002.
- Rosenschein S.J., Kaelbling L.P.: A Situated View of Representation and Control, in Agre P.E. and Rosenschein S.J. (eds.), *Computational Theories of Interaction and Agency*, MIT Press, Cambridge/Boston/London, pp.515-540, 1996.
- Salovey P.: Introduction: Emotion and Social Processes, in Davidson R.J. et al. (eds.), *Handbook of Affective Sciences*, Oxford University Press, Oxford New York, pp.747-751 (Chapter 38), 2003.
- Scherer K.R.: Emotion as a process: Function, origin, and regulation, *Social Science Information*, 21:555-570, 1982.
- Segalowitz S.J., Cohen H.: Editorial: Affective Neuroscience, Brain and Cognition, *Affective Neuroscience*, 52(1), p.1, 2003.
- Solomon R.C.: President's Column: "Defining our Terms", *Emotion researchER*, Official Newsletter of the International Society for Research on Emotion, 15(2), pp.3-4,6, 2001.
- Wegner P.: Why Interaction is More Powerful Than Algorithms, *Communications of the ACM*, 40(5), 80-91, 1997.
- Wehrle T., Scherer K.R.: Potential Pitfalls in Computational Modeling of Appraisal Processes: A Reply to Chwelow & Oatley, *Cognition and Emotion*, 9(6):599-616, 1995.
- Weizenbaum J.: ELIZA - A Computer Program for the Study of Natural Language Communication Between Man and Machine, *Communications of the ACM*, 9(1):36-45, 1966.