Emotional Agents need Formal Models of Emotion

Joost Broekens and Doug DeGroot

Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands, Email: {broekens,degroot}@liacs.nl

Abstract

Embedding a computational model of emotion in virtual agents is beneficial in a variety of domains. These domains include gaming, VR training, HCI and electronic tutors. Although these domains have different motives for embedding such a model, they share the same overall approach. Once the requirements for the agent are clear, first an emotion theory is chosen as basis for the computational model. Second, the model is implemented and embedded into the virtual agent. Candidate emotion theories are mostly cognitive, i.e. appraisal theories. Furthermore, theories used for this purpose are mostly structural descriptions -usually represented by text and tables- of the relations between events, evaluations of events and emotions. Structural descriptions are abstract, but computational models are concrete. In this paper we explain the nature of the gap between the level of abstraction of these structural descriptions of appraisal theories and computational models of emotion. We also show that this gap introduces several important problems that make it hard to evaluate the consistency between a computational model of emotion and the appraisal theory it is based on. Lastly, we propose a formalism to narrow this gap, which can be used to describe the structure of appraisal. We believe that our formalism stimulates the consistency of computational models based on appraisal theories and thereby increases the potential and plausibility of emotions in virtual agents and robots.

1. Introduction

In cognitive psychology, emotion is often defined as a psychological state or process that functions in the management of goals, desires, concerns and needs (we refer to these four terms as *goals*). According to this definition, this state consists of physiological changes, feelings, expressive behaviour and inclinations to act. Emotion is elicited by the evaluation of an event as positive or negative for the accomplishment of the agent's goals. Thus, an emotion is a heuristic that relates the events from the environment to the agent's goals [7]. Additionally, emotions are used in non-verbal communication. Inspired by this heuristic and communicative aspect of emotions, computational models of emotion are embedded in virtual agents in a variety of domains, including:

- HCI and electronic tutors: emotions are embedded primarily because they can be used as intuitive communication medium to better understand the human, to act upon this understanding accordingly, or to express the state of the tutor [4].
- Games: emotions are embedded in the non-player-characters for entertainment and realism purposes. The communicative aspect of emotional expression is used to create a sense of realism [6].

- Virtual-reality safety-training environments: agents are embedded with emotions primarily to create an enhanced sense of realism for the trainees, through the emotional expression of the virtual agents in the training [3]. It is hoped that this will increase training efficiency, resulting in fewer accidents.
- Decision-making and planing: there is a large body of literature on the embedding of emotions in this area. It is more and more recognised that emotional models can be used as useful heuristic for the construction and evaluation of plans [2][3].

The majority of computational models of emotion embedded into virtual agents are based on appraisal theories, cognitive theories of emotion that attempt to explain why a certain event results in one emotional response rather than another and why a certain emotion can be elicited by different events. The key concept of appraisal theories is that the subjective evaluation of the environment in relation to the agent's goals is responsible for emotions [9]. This evaluation is called appraisal. Appraisal theories contrast with, for example, the James-Lange theory. In this theory, an emotion is the interpretation of the bodily reactions that are provoked by an event, while appraisal theories assume that bodily reactions are a result of the emotion, which is a result of cognitive evaluation. Appraisal theory also contrasts with the Schacter-Singer cognitive theory. In this theory, an emotion results from the cognitive evaluation labelling the arousal of the organism. Arousal results directly from events, and cognitive processing is not needed for this. Thus, the emotion is based on arousal and differentiated by cognitive evaluation, while appraisal theory assumes that arousal itself is a result of cognitive evaluation. In short, appraisal theory focuses on emotion being a result of the cognitive evaluation of the environment in relation to the agent's goals (desires, concerns and needs), which explains its popularity in computational models of emotion.

Typically, appraisal theories that are used for computational models of emotion are descriptions of the relations between events, appraisal of events, and emotions. Such descriptions are abstract, but computational models are concrete. In this paper we explain the nature of the gap between the level of abstraction of these structural descriptions of appraisal theories and computational models of emotion. We also show that this gap introduces several problems that make it hard to evaluate the consistency between a computational model of emotion and the appraisal theory it is based on. Lastly, we propose a formalism to narrow this gap, which can be used to describe the structure of appraisal.

2. Structure, process and computation

One common classification of appraisal theories is based on structural versus processual description [9]. Structural theories of appraisal (also called "black-box models" or "structural models") describe the structural relations between (1) the environment of an agent, (2) the agent's appraisal functions that interpret the environment in terms of values on a set of subjective measures, called *appraisal dimensions*[1] and (3) the functions that relate these values to the agent's emotions. Process theories of appraisal describe, in detail, the cognitive operations, mechanisms and dynamics by which the

[1] An appraisal dimension is a variable - e.g., agency or valence -, used to express the result of the appraisal of a perceived object - e.g., a friend - that influences emotion.

appraisals, as described by the structural theory, are made [11]. From a computational (or cognitive) point of view, a structural theory of appraisal thus aims at describing the declarative semantics of appraisal, while a process theory of appraisal complements this description with procedural semantics.

Computational models resemble, but differ from process theories of appraisal. On the one hand, they both involve detailed operations. Computational models involve operations that control a "Turing machine" device, while process models involve operations that control a "cognitive device". On the other hand, process models of appraisal in the literature are seldom detailed enough, or even suitable, to be directly implemented as a computational model. In fact, detailed cognitive operations are rarely (if ever) algorithmically described. As a consequence computational models of emotion are often inspired by structural theories of appraisal. This relation between structural theory and computational model is graphically represented in Figure 1 by the dotted arrow. A prototypical model based on a structural theory of appraisal can be found in [8]. In this paper, we use *computational model* when referring to computational models of emotion and *structural theory* when referring to structural theories of appraisal.

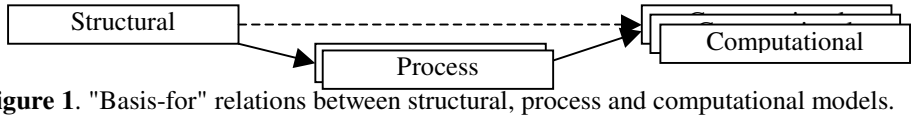

Figure 1. "Basis-for" relations between structural, process and computational models.

3. Embedding computational models of emotion

Developing a computational model of emotion for a virtual agent involves several steps. In general these steps are:

- Deciding what an emotion is and how it is used.
- Within a domain, the requirements for the virtual agent in need of an emotional model are defined. These requirements constrain the set of possible theories that are suitable for the model.
- A suitable structural theory is chosen and used as basis for a computational model.
- The model is designed and built, and finally embedded in the agent and then tested.

Developing a computational model based on a structural theory involves making a number of assumptions that relate to computational aspects like timing, dependency between appraisal functions, priority of appraisal functions, possible values of appraisal dimensions and event categories related to appraisal functions, and so on. When an event occurs in the environment, a computational model needs information about when/if this event is appraised, which appraisal functions are involved, how long this evaluation takes, which appraisal functions are responsible for the resulting emotion and how these functions are related to one another. Systematic psychological study has only recently started to give answers to these questions [10]. However, from a computational point of view these answers still lack many details. Developing a computational model still needs a number of assumptions that relate to computational aspects. Also, the answers are often presented informally, which easily introduces interpretative errors. This lack of formality and detail, needed as basis for computational models, is what we call the *representational gap* between structural theories and computational models. Several important problems arise from this gap:

- Inconsistency between computational model and structural theory. The theory and the model might or might not be consistent, due to the many assumptions and interpretations needed to develop a computational model.
- Problematic identification of bugs versus features. If a computational model is developed and subsequently tested in a virtual agent, the resulting emotions can be different than what was expected. Since there is no specification of the structural theory that is suitable for the model, there is no guideline for the correct interpretation of a phenomenon as bug or not. This could be a major problem, if it results in 'tweaking' the model, while actually the theory should be 'tweaked'.
- Incorrect interpretation of all of the ramifications of a structural theory in an early stage of development. Switching to a second theory - when the initial theory appeared to be unsuitable - consumes precious development time and effort.
- Lack of an implementation-independent formal description of a structural theory. This has at least three drawbacks. First, the computational interpretations have to be made over and over again, which is a loss of intellectual effort. Second, small but potentially important changes to the theory can remain unnoticed for a long time. Third, there is no way to compare one model with another both implementing the same theory, while this could be very useful to identify potential inconsistencies in the theory.

4. Attempting to narrow the representational gap

In a first attempt to narrow this representational gap we have developed a formal notational scheme to specify the declarative semantics of a structural theory of appraisal. Our formalism is built around sets of perception processes, appraisal processes and mediating processes (Figure 2). The notation used for these three types of processes and the accompanying terminology have been adopted from [11]. We now briefly describe the components of the notation. The external world, W, is the set of all events that can occur in, and objects that can reside in the environment. Perception processes - the set P - filter, select and translate information from the external world, and produce *mental objects* - representations of the external world suitable for appraisal. We understand the set of mental objects - the set O - produced by the perception processes as the current content of working memory. Appraisal processes - the set A - evaluate the mental objects produced by the perception processes and assign appraisal dimension values - represented by the set V - to these objects. Some appraisal processes may be relevant to emotion only through their influence on other appraisal processes. In this case these "indirect" appraisal processes assign only zero-values to evaluated mental objects. Mediating processes relate appraisal information to emotions. Thus, mediating processes - the set M - relate appraisal dimension values to emotion-component intensities - the set I -.

The formalism also allows specification of perception processes that perceive the agent's current appraisal dimension values and current emotion components. These two kinds of information are translated to mental objects. Since only perception processes put information in working memory, this means that in our formalism emotion-component intensities - the set I - and appraisal information - the set V - must be perceived before the agent is able to use these two kinds of information as mental

objects in appraisal. Separating conscious emotional influence - from V and I to P - from unconscious emotional influence - from I to A - allows specification of appraisal processes that are biased by a specific combination of emotional feedback (i.e. no feedback, unconscious, conscious, both). This allows, for example, explicit specification of appraisal processes involved in coping, re-appraisal and strategic use of emotions.

To specify the structure of the set of perception, appraisal and mediating processes, our formalism allows the specification of process-dependencies. For example, some process-dependencies can be defined as exhibitory relations, while others can be defined as inhibitory relations between processes.

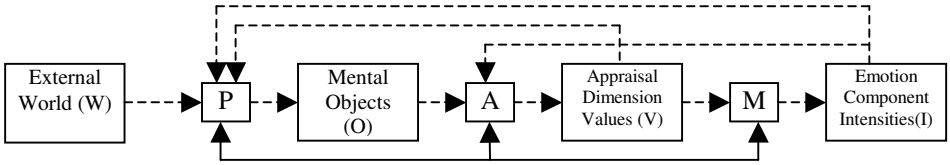

Figure 2. A graphical overview of the components of the formal notation. Dotted arrows denote input for processes, while solid arrows denote potential process dependencies.

4.1. Formal Notation

Please note that in this paper n is used as arbitrary number to denote multiplicity. Two sets both with n elements do not necessarily have the same number of elements. When two sets do have the same number of elements another subscript is used, e.g. m.

World, perception processes and objects of appraisal.

Definition 1.1: $W=\{w_1,\ldots,w_n\}$ is the set of all observable objects and events in the environment of the agent.

Definition 1.2: $O=\{o_1,\ldots,o_n\}$ is the current content of working memory and is the set of all mental objects currently perceived by the agent, with $o_i=(t,$ object_name$)$ and $t\in OT$, the set of mental object types as defined next.

Definition 1.3: $OT=\{t_1,\ldots,t_n\}$ is the set of type names - (O)bject (T)ypes - used to specify mental object types (e.g. belief, goal, like, dislike, stimulus modality type, etc).

Definition 1.4: If we define V as the set of appraisal dimension values (see definition 2.2) and I as the set of emotion-component intensities (see definition 3.2) then $P=\{p_1,\ldots,p_m\}$ is the set of all perception processes available to the agent, with $p_i: W^n\times V^n\times I^n\rightarrow O^n_i$. A perception process p_i thus selects and translates one or more objects and events in the agent's environment W, the agent's current appraisal dimension values V and it's emotion-component intensities I, to a subset O_i of mental objects O. The set P maps $W^n\times V^n\times I^n$ onto the set O, resulting in $O_1\cup\ldots\cup O_m=O$. Perception processes are intimately linked with attention.

Appraisal processes, appraisal values and dimensions.

Definition 2.1 $D=\{d_1,\ldots,d_n\}$ is the set of appraisal dimensions, containing elements like agency and valence.

Definition 2.2: $V=\{v_1,\ldots,v_n\}$ is the set of appraisal dimension values attributed to mental objects, with v_n equal to a one-dimensional value resulting from the appraisal of one or more mental objects. $V\subseteq O^n\times D\times[-1,1]$ with D the set of dimensions, O the set of

mental objects to which the appraisal dimension value is attributed, and $[-1,1]$ the set of real numbers representing possible values.

Definition 2.3: $A=\{a_1,\ldots,a_m\}$ with $a_i: O^n \times I^n \rightarrow V_i^n$, a_i is an appraisal process, mapping mental objects to elements from the appraisal-process-specific subset $V_i \subseteq V$ of possible appraisal dimension values. V_i is the appraisal-result of appraisal process i, and $V_1 \cup \ldots \cup V_m = V$. Appraisal can be unconsciously biased by the current emotion, explaining I^n as input for the appraisal processes.

Formalising the mediating processes in R.

Definition 3.1: $E=\{e_1,\ldots,e_n\}$ is the set of emotion-components, like subjective feelings, specific facial expressions, physiological reactions and action tendencies.

Definition 3.2: $I=\{i_1,\ldots,i_n\}$ is the set of emotion-component intensities. $I \subseteq E \times [0,1]$ with $[0,1]$ the set of real numbers representing the possible intensity, and E as defined above.

Definition 3.3: $M=\{m_1,\ldots,m_m\}$ are processes that mediate between appraisal dimension values and emotion-component intensities. $m_j: V^n \rightarrow I_j$ is a mediating process typically mapping n elements from the set V of appraisal dimension values to a subset of emotion-component intensities $I_j \subseteq I$, with $I_1 \cup \ldots \cup I_m = I$.

Process dependency and data constraints. Our formalism facilitates the representation of the structure of processes using guarded process-dependencies. To be able to define the notation for process-dependencies, we first define guards and dependency types.

Definition 4.1: The set $G=\{g_1,\ldots,g_n\}$ of guards is the set of second-order predicates over the elements of the sets P, O, A, D, V, M, E and I, and over the variable i, being the actual value of elements in the set V and the intensity of the emotion-components in the set I. This allows definition of conditional dependencies between processes.

Definition 4.2: The set $LT=\{n_1,\ldots,n_n\}$ is the set of dependency type names - (L)ink (T)ypes - used to identify the nature of the dependency between two processes (e.g. inhibitory, causal, correlation, information flow, parallelism, etc).

Definition 4.3: Let L be the set $L=\{l_1,\ldots,l_n\}$ with $L \subseteq PP \times PP \times G \times N$ and $PP=P \cup A \cup M$. The elements of L define dependencies - (L)inks - between processes constrained by: $(\forall x)(\exists y)$ processing in q_x is influenced $\mathtt{iff}$ $((p_y,q_x,g,n) \in L \wedge g=\mathtt{true} \wedge p_y,q_x \in PP \wedge g \in G \wedge n \in N)$. If a dependency exists between a process p_y and q_x and the guard of that link is true, processing in q_x is influenced in a way denoted by the type n.

Definition 5.1: The set $H=\{h_1,\ldots,h_n\}$ is the set of data constraints and is defined as a set of second-order predicates just like the set G. These data constraints are global, and not attached to process-dependencies. They enable specification of relations between data that must hold according to the structural theory.

5. Discussion

Several extensions to the formalism (notably to O and P) can be found in [1], including the rational for the different elements of the formalism approached from an appraisal theoretic perspective and several examples of how to use the formalism. In this paper we

briefly discuss one of these extensions, i.e. time, and explain how a formalism like ours can actually stimulate the consistency of computational models of emotion in virtual agents.

Time. We believe the current version of our formalism is a good first attempt to narrow the representational gap, however the aspect of time is still missing. All sets in the formalism are static sets. A formal model based on such timeless sets essentially represents the appraisal-structure of the mind of an agent at either one instant or at all possible instants in time (whatever suits best). If a computational model addresses appraisal processes, including detailed aspects of environment evaluation and emotional responses, and if we want to be certain that a computational model is consistent with the structural theory it represents, then even a structural theory needs to consider time, and a formalism dealing with structure thus needs to be able to represent time (see also [12] for a comparable argument). In [1] we address this issue in detail, and we re-define all sets that are used in our formalism as timed sets, by slicing them up in timed sub-sets. This, in combination with *add* and *remove* operations on these timed sub-sets, allows for a formal description of a large number of phenomena including developmental changes to the appraisal structure of the agent, detailed causal relations between processes and emotion reaction-time.

Theory and data exploration. With the addition of time, our formalism can be used to specify both static structures of appraisal (i.e. a structural theory) and dynamic experimental results. For example, at time t=0, the data sets `O`, `V` and `I` can be specified to contain certain elements, while after the manipulation of the human subject (say time t=10) the sets can be specified to have different elements, based on experimental results. Formal modelling of both experimental results and structural appraisal theory using the same formalism greatly facilitates four things. First, it allows detailed comparison of theoretical predictions with experimental results using the same formal notation, and second, it facilitates implementation of these results in a computational model. Third, it facilitates the construction of a single formal database of results obtained from experiments with human subjects (c.f. [12]) and results obtained from experiments with computational models. Such a database can be used to evaluate the results generated by new experiments with computational models. Fourth, appraisal theorists do many experiments to test their theories, and the results are hard to keep up with for computer scientists. Online formal databases that contain human-subject results, computational results, and different structural theories can help to enhance the consistency of computational models of emotion by, for example, automated theory-change tracking and automated experimental-result consistency checks.

6. Conclusion

We have argued that formal models of structural appraisal theories are necessary for the development of consistent computational models of emotion. We have presented a formalism that attempts to narrow the gap between structural appraisal theories and computational models of emotion. The formalism can be used to specify the declarative semantics of a structural theory of appraisal, and it can be used to formalise

experimental results obtained with computational models of emotion or human subjects. This stimulates the consistency of computational models based on appraisal theories and thereby increases the potential and plausibility of emotions in virtual agents and robots.

References

[1] J. Broekens and D. DeGroot. *Formal Models of Emotion: Theory, Specification, and Computational Model*. LIACS Technical Report, 2004.

[2] A. Coddington and M. Luck. Towards Motivation-based Plan Evaluation. In Russell, I. and Haller, S., Eds. *Proceedings of Sixteenth International FLAIRS Conference*, pages 298-302, Florida, USA, 2003.

[3] S. Marsella and J. Gratch. Modeling the Interplay of Emotions and Plans in Multi-Agent Simulations. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, Edinburgh, Scotland, 2001.

[4] D. Heylen, A. Nijholt, R. op den Akker and M. Vissers: Socially Intelligent Tutor Agents. *IVA 2003*: 341-347, 2003.

[5] M. D. Lewis. Personal Pathways in the Development of Appraisal. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, Methods, Research*, Oxford University Press, 2001.

[6] B. Mac Namee and P. Cunningham. Creating Socially Interactive Non Player Characters: The m-SIC System. *IJIGS* 2(1), 2003.

[7] K. Oatley. Emotions. *The MIT Encyclopedia of the Cognitive Sciences (MITECS)*. Edited by Robert A. Wilson and Frank Keil, ISBN 0-262-23200-6, 1999.

[8] A. Ortony, G.L. Clore and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

[9] I.J. Roseman and C.A. Smith. Appraisal Theories: Overview, Assumptions, Varieties, Controversies. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, Methods, Research*, Oxford University Press, 2001.

[10] C. van Reekum. *Levels of Processing in Appraisal: Evidence from Computer Game Generated Emotions*. Phd Thesis nr. 289, University de Geneve, Section de Psychology, 2000

[11] R. Reisenzein. Appraisal Processes Conceptualized from a Schema-Theoretic Perspective: Contributions to a Process Analysis of Emotions. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, Methods, Research*, Oxford University Press, 2001.

[12] T. Wherle and K.R. Scherer. Towards Computational Modeling of Appraisal Theories. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, Methods, Research*, Oxford University Press, 2001.