# Formalizing Cognitive Appraisal: From Theory to Computation

**Joost Broekens and Doug DeGroot**
Leiden Institute of Advanced Computer Science, Leiden University
The Netherlands
broekens@liacs.nl, degroot@acm.org

## Abstract

Cognitive appraisal theory has been used extensively as basis for computational models. Evaluating these models in an objective way is a difficult issue. Further, integration and comparison between appraisal theories is needed in order to advance the development of appraisal theory. In this paper we argue that a good solution is to formally specify structural appraisal theories using a common formalism. Applying such a formalism results in an objective representation of the structural appraisal theory. We provide evidence for the claims that such a representation (1) can be used to evaluate the computational model, and (2) helps to advance the development of appraisal theory as it can provide a well-structured representation of the theory that can be used to compare and merge theories, as well as make theoretical predictions explicit.

## 1 Introduction

Computational models of emotion are used in a wide variety of artificial emotional agents including electronic tutors [Heylen et al., 2003], social robots [Breazeal, 2001], virtual agents in VR training environments [Henninger et al., 2002], and agents targeted at decision-making and planning [Coddington & Luck, 2003]. Many of these models are based on a cognitive appraisal theory. Such theories explain human emotions as a result of the subjective evaluation of events that occur in the environment. However, these theories lack the necessary detail to base a computational model upon [Gratch & Marcella, 2004]. As a result, it is difficult to evaluate if such a model correctly implements the theory. Further, to advance the field of appraisal theory, theories need to be integrated and compared [Wehrle & Scherer, 2001]. This paper has two threads; we argue why a common formal notation for the structure of appraisal is needed to (1) build and evaluate a computational model of emotion and (2) advance appraisal theory. We also propose a particular notation and provide evidence of its usefulness.

## 2 Structure, process, computation

First, we analyze the relation between cognitive appraisal theory and computation. We argue that it is useful to have a theory-independent formal notation to describe structural appraisal theories (i.e., the behavior of processes that play a role in appraising a situation, how these processes are linked to each other, what the resulting emotions could be, etc.).

A common classification of appraisal theories is based on a structural versus a process-based description [Roseman & Smith, 2001]. Structural theories of appraisal (also called black-box models or structural models) describe the structural relations between:

- the environment of an agent and perception of this environment: *perception*;
- the agent's appraisal processes that interpret the perceived environment in terms of values on a set of subjective measures, called appraisal dimensions[1]: *appraisal*;
- the processes that relate these values to the agent's emotions: *mediation*.

Process theories of appraisal describe, in detail, the cognitive operations, mechanisms and dynamics by which the appraisals, as described by the structural theory, are made and how appraisal processes interact [Reisenzein, 2001]. In other words, a structural theory of appraisal aims at describing the declarative semantics of appraisal, while a process theory of appraisal complements this description with procedural semantics. In this paper we adopt the terms structural model (SM) and process model (PM) respectively, and use appraisal theory/model when referring to cognitive theories/models of appraisal in general.

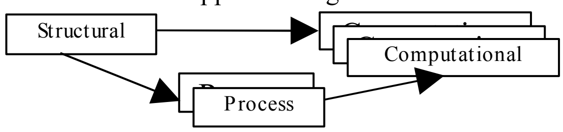


Figure 1. Three possible translations between structural, process and computational models of emotion.

A computational model is a model that is composed of operations that unambiguously control the behavior of a device. These operations may use available input data. If there is a sequence of such operations that

[1] An appraisal dimension influences emotion and can be considered as a variable—e.g., agency or valence—, used to express the result of the appraisal of a perceived object—e.g., a friend.

maps a specific input to a specific behavior (output), an algorithm exists for that mapping. In this paper, we define a computational model (of appraisal) (CM) as a structured collection of interacting algorithms that operate serially or in parallel, with operations that are eventually reducible to the Turing machine level.

In general, there is a generic-to-specific relation between structural, process and computational models. SMs are the basis of computational and process models, and PMs are also the basis of CMs. In this case 'basis of' usually means that a model *A* that is the basis of a model *B* contains less details than model *A*, and therefore different model *B* instantiations are possible based on model *A* (Figure 1). Although this is true in general, in this section we first argue that the difference between a structural, process and computational description is not only of *degrees of detail* but also of *kind*; all three models are equally important for appraisal theory. We then argue that a formal description of the structural model is needed for consistency and evaluation of a CM of emotion. Finally we argue that such a description is also needed for the advancement of appraisal theory.

## 2.1. Computational models are not detailed process models

A CM resembles, but differs from a PM. Both involve operations; the first involve operations that control a Turing machine device, while the second involve operations that control a 'cognitive device'. However, PMs in the literature are seldom detailed enough, or even suitable, to be directly implemented as a CM. Cognitive operations are rarely algorithmically described, and lack constraints that are important for a CM. Consequently, CMs are often inspired by structural models of appraisal. CMs are implemented using mechanisms borrowed from artificial intelligence research [Gratch & Marsella, 2004]. This results in a large gap between the SM and the CM. To narrow this gap two approaches seem plausible. First, develop a common formal notation to describe SMs, as proposed in this paper. Second, enhance PMs so that they include enough detail to serve as CMs. For the latter, appraisal theorists need to computationally describe PMs [Reisenzein, 2001]. We argue that, *from a system-engineering perspective*, enhancing process models to the level of computational model is not the preferred solution and that structural, process *and* computational models are needed and complementary.

To argue for this point we use the software architecture paradigm as a metaphor for the relation between SM, PM and CM. A system—a collection of components organized to accomplish a specific function or set of functions; in our case the appraisal system—can be described using an architectural description, or architecture [IEEE, 2000]. An architecture is composed of different architectural views. A view is a representation of a whole system from the perspective of a related set of concerns. Each view has its own set of concerns, associated notation, modeling and analysis techniques, all of which are defined by a viewpoint [IEEE, 2000], the glasses through which one looks at a system in order to come up with a suitable description of that system addressing a specific set of concerns. Taking appraisal theory as an example, one can use a process-oriented viewpoint as basis for a PM (view) of appraisal. Alternatively one can take a structural viewpoint as the basis for a SM (view) of appraisal.

First, all software development effort should start with the specification of the *system requirements*. In general these requirements specify the functions and behaviors of the system using natural language, i.e., what the system should do.

Second, the *structural view* specifies the different kinds of data, the components that are responsible manipulating these different kinds of data, and the relations between components, kinds of data or both. A SM of appraisal corresponds best to the structural view. In an appraisal context, SMs specify the 'what', the relations between appraisal processes, appraisal dimensions and emotions.

Third, the *process view* specifies how the interaction of processes is responsible for the production and interpretation of data and how the data flows through sequences of processes. This view typically includes time as well as a description of the dynamics of the system. This view corresponds to a PM of appraisal using cognitive operations to describe how appraisals evolve and how emotion results from these appraisals.

Fourth, the *computational view* provides a functional decomposition of the system into objects that interact at interfaces [IEEE, 2000]. It organizes the operations that implement the processes in the process view into logical clusters that cooperate, often called packages, classes and components (i.e., a structured collection of interacting algorithms). These packages and classes specify the exact operations needed to effectuate the processes specified by the process view, using data, constraints, and relations as specified by the structural view. This view corresponds to the CM, unambiguously specifying the operations that implement the manipulations in the processes defined by the PM, according to the SM's constraints.

Finally, the *engineering* and *technology views* specify the mechanisms and hardware required to support and implement the computations defined by the computational view. In natural agents, these views would correspond to descriptions of the signal processing in and interaction between the different structures, lobes, gyri and nuclei of the brain and the available senses. In artificial agents, these views correspond to the computer's architecture, processors, available input/output devices, etc. The more the artificial agent's hardware (or simulated hardware, using, e.g., cognitive modeling packages like PDP++, ACT-R, Soar, etc) resembles the natural agent's hardware, the more the computational view of the artificial agent resembles that of the natural agent.

The relation between the different views corresponds to the relation between the structural, process and computational models of emotion. Views are complementary in the sense that each one specifies necessary aspects of the system and that one specific view is non-reducible to another, because of their different viewpoints. Some views describe processes; others describe structure while again others describe the computational components.

Both structural and process views are necessary for 'an architecture' of appraisal. This has been acknowledged by appraisal theorists, stating that both structural theories as well as process oriented theories are needed for the development of appraisal theory [Roseman & Smith, 2001]. Although it has been argued that if cognition is computation then PMs should eventually be CMs [Reisenzein, 2001], our system-engineering-inspired analysis shows that this is not necessarily the case. CMs involve more detail, but are also of a different kind than PMs. We conclude that for the development of appraisal theory, we need structural, process and computational models to provide complementary descriptions of the system of appraisal[2].

### 2.2 Consistency and evaluation of computational models of appraisal

Development of CMs is greatly facilitated by a formal declarative description of the intended behavior of the components of the model. Specifications facilitate, e.g., discussion about the assumptions and premises of a computational model, comparison between different CMs and, identification of bugs in the CM versus inconsistencies in the specification. The importance of formal specifications is generally acknowledged [IEEE, 2000] and the main cause for the emergence and subsequent success of research into the areas of software architecture and specification.

Currently, appraisal theories are described in a way that is insufficiently precise as a specification for a CM of emotion [Gratch & Marsella, 2004]. An important consequence is that SMs are used as basis and that many assumptions are needed to develop the algorithms that model what a theory describes [cf. the 'coding problem' as described by Mallery; 1988]. We argue that a common formal notation to describe an underlying SM facilitates the development of a CM.

Additionally, we argue that a formal description is critical to evaluate the consistency of these computational models. Currently, such models are evaluated for consistency using, for example, extensive scenario-testing [Gratch & Marsella, 2004]. These scenarios are first presented to human subjects and then presented to the computational model. Results are compared. Most models, however, are evaluated based on 'face-value', i.e., 'if the emotional behavior looks good, the model must be ok'. The real problems are 'what if it doesn't look good', and who defines 'good'?

Although 'face-value' and similar approaches [cf. Henninger et al., 2003] to model-evaluation give important insights regarding believability and consistency of the behavior of the computational model, these approaches are insufficient to evaluate the consistency of the CM for the following reasons.

First, comparisons between model and real-world are essentially based on results from the computational model versus human behavioral results (actions, facial expressions, introspective reports, etc.). However, behavioral data may not be a good representation of what is actually happening to the appraisal processes in the human subject. This is a fundamental problem; a computational agent's unexpected emotion cannot be discarded as being the result of a bug in the computational model. It is unclear if the human's emotion to which the agent's emotion is compared was influenced by an unknown factor. There is no objective model against which to compare both emotions. Without a formal description of the SM, there is no specification of the appraisal theory on which the CM is based. Therefore, there is no formal way of identifying whether the model or the theory should be changed in case of a discrepancy between the human subjects and the computational model [cf. 'mechanism artifacts' as described by Mallery; 1988]. This could result in tweaking a CM to achieve the desired behavioral result, especially as many assumptions are needed before a SM can be used as basis for CM. Therefore, evaluation of consistency can not rely solely on behavior-based evaluation. An objective model is needed against which to compare the human and computer model's emotions.

Second, appraisal theorists are specialized in the evaluation of appraisal models, and already have a large body of results that include physiological, introspective and behavioral data. Many of such experiments are being done and a large amount of work goes into the interpretation of the results. Verification of a small part of a theory takes a large amount of effort (see, e.g., van Reekum [2000]). This points to the necessity of reuse of this information for effective evaluation of computational models of emotion. A *common* formalism can help to structure the information produced by these experiments and promote reuse.

To summarize, if we want to evaluate the consistency of CMs in artificial agents then two issues are critical. First, before designing CMs at the algorithmic level, declarative information is needed on the processes that are responsible for perception, appraisal and mediation as defined by the appraisal theory. Second, objective information is needed to evaluate the consistency between CMs and appraisal theory, and reuse of experimental findings is essential.

### 2.3 Formal specification is needed to debug emotional agents

In the previous section we have talked about evaluating and debugging computational models using a formal notation of the structural model. It could be argued (although *we* don't) that formalization of the behavior of emotion processes is *not* needed to find inconsistencies in a CM. Essentially, such CM is a collection of algorithms, and it should be possible to debug these algorithms and find exactly what is wrong.

[2] Thus we agree with the essence of Reisenzein's [2001] previous argument: to really understand emotion (from a cognitive appraisal perspective) we need to get down to the computational details. To extend this argument we suggest that *if* cognition is computation then software engineering techniques are useful for the development/analysis of cognitive theories.

Following this line of reason, it seems that in order to debug a CM, a formalism that describes what an emotion is in terms of beliefs, desires and intentions (BDI) is needed the most, as such formal description permits to predict the resulting emotion and debug the agent's BDI system. To the contrary, we argue that in the near future, formalization at the structural level is as necessary as BDI-based formalization of emotion.

Artificial agents are getting more and more complex. Consider agents like CMattie developed by the group of Stan Franklin already in 1999 [Franklin & Graesser, 1999], or the emotional robot Kismet [Breazeal, 2001]. These agents are based on an amalgam of psychological theories and composed of many different components (including models of emotion). These agents learn, adapt, and develop, and the results produced by this research are published in the psychological literature [Baars & Franklin, 2003].

In the future, identifying what the emotions, but more importantly the exact desires and beliefs, of these agents are, will be become a larger and larger problem, at least from a "classical debugging" point of view (i.e., tracing execution of a CM). We might have to resort to simply asking the agent how it feels and what it thinks about a certain topic. However, this results in a subjective report of the artificial agent's internal state. If this report is incompatible with our own feeling of the situation, can we just discredit this subjective report based on our own feeling and list it as a bug to be solved? How are we to decide that this specific agent's architecture indeed has a problem resulting in such-and-such emotion? We think we will not always be able to debug such artificial agents.

We need a declarative description of the processes that are responsible for an agent's emotion, in order to evaluate if the agent's unexpected emotion resulting from an experimental situation is due to a problem in the agent's architecture, or due to a mismatch between our interpretation of the situation and the agent's interpretation. That is, are we—evaluating observers—wrong or is the agent's architecture wrong (or is the predicted emotion wrong and have we found something fundamentally new about emotions)?

The relation between a BDI-based formalism and a formal notation of a structural model as proposed here could be that the former specifies how specific appraisals result from the interpretation of a situation, while the latter specifies the behavior of appraisal and emotion processes at a meta-level (i.e., minimum and maximum level of activity of processes, potential activation propagation through processes, see 2.5).

### 2.4 Formal specification is needed to advance appraisal theory

Apart from the problem of translating appraisal theories into computational models, another problem exists that is more directly related to appraisal theory. There are many different appraisal theories [Frijda & Mesquita, 2000; Ortony, Clore & Collins, 1988; Reisenzein, 2000; Scherer, 2001; Smith & Kirby, 2000]. Wehrle & Scherer [2001] argue that to advance appraisal theory development, comparison between and integration and convergence of theories is necessary, and that this is facilitated by a formal description of a theory as this description allows detailed predictions, and precise and structured revision of a theory.

A second reason to formalize a theory is that the process of formalization clarifies and refines theories [Mallery, 1988]. Formalization requires a researcher to commit to specific definitions of components and interaction between these components as hypothesized by a theory. This commitment clarifies the actual structure of the theory and refines the consistency of the theory, even though the resulting formal model might not have predictive power [Mallery, 1988].

### 2.5 Formalism Requirements

Typically, a common formal notation should allow the description of a SM such that this description includes the following data (of which many are also relevant to PMs [Reisenzein, 2001]):

- the nature and level [van Reekum, 2000] of processes; deliberative, automatic, innate?
- The relation between (results of) perception and appraisal processes.
- When and how are these processes activated? Are there thresholds? Can activation be sub-threshold?
- What kind of input and output (representations) does a certain process needs/produce?
- Does a process continuously output results or periodically (how often)?
- How many and what perception, appraisal and mediating processes exist?
- Is information activation binary or gradual? E.g., how strongly must a certain event be perceived for it to be input for a certain appraisal process?
- the number of different appraisal dimensions, their activation range and the responsible processes.

## 3 A set-based formalism for the structure of appraisal

In this section we briefly introduce the basic concepts of the first version of the formal notation we propose to describe structural theories of appraisal (for more detail see [Broekens et al., *submitted*; Broekens & DeGroot, 2004]. Our formalism is set-based and 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 are borrowed from Reisenzein [2001]. The external world, W, is the set of all events and objects that can respectively occur and 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 define the set of mental objects produced by the perception processes, the set O, as the current content of working memory. Appraisal processes, the set A, evaluate the mental objects produced by the perception processes and assign a combination of appraisal dimension values, the set V, to these 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.

Perception processes also perceive the agent's current appraisal dimension values and current emotion components. These two kinds of information are translated into mental objects. Since, by choice, in our notation only perception processes can put information into working memory, the emotion-component intensities, I, and appraisal information, V, must be perceived before the agent is able to use these two kinds of information in appraisal. This is consistent with the idea that appraisal (viewed as a cognitive evaluation process) is the result of manipulation and evaluation of objects in working memory. Additionally, separating conscious emotional information—i.e., V and I perceived by P—from unconscious emotional influence—I influencing A—allows the specification of appraisal processes that are biased by a specific combination of emotional feedback (i.e., none, unconscious, conscious, both). This enables explicit specification of appraisal processes involved in coping, re-appraisal and strategic use of emotions. This ability is important for the completeness of our formalism.

To describe the *structural relations* between elements in the sets of perception, appraisal and mediating processes, our formalism allows the specification of process-dependencies in the form of second-order predicates. For example, process-dependencies can be defined as excitatory relations or inhibitory relations between processes.

We stress that our goal is to propose the use of a common formal notation, not a theory of appraisal. Such notation is necessarily based upon appraisal theoretical assumptions. However, other assumptions can be taken as basis if the current ones are too restricting (e.g., too strict a separation between P, A and M).

## 4 Facilitation of theory integration

In the context of appraisal theory, important criteria of success for a formal notation describing the structure of appraisal are to what extent the notation facilitates comparison, convergence and integration [Wehrle & Scherer, 2001]. We use our formalism to formalize a potential integration of Scherer's [2001] Stimulus Evaluation Checks (SEC) model and Smith & Kirby's [2000] Appraisal Detector Model (ADM) process model. We call this model the SSK model (*S*cherer-, *S*mith and *K*irby). Obviously, the SSK model should not be interpreted as a complete integration, as our goal was to show the utility of a common formal notation in the emotion theory domain.

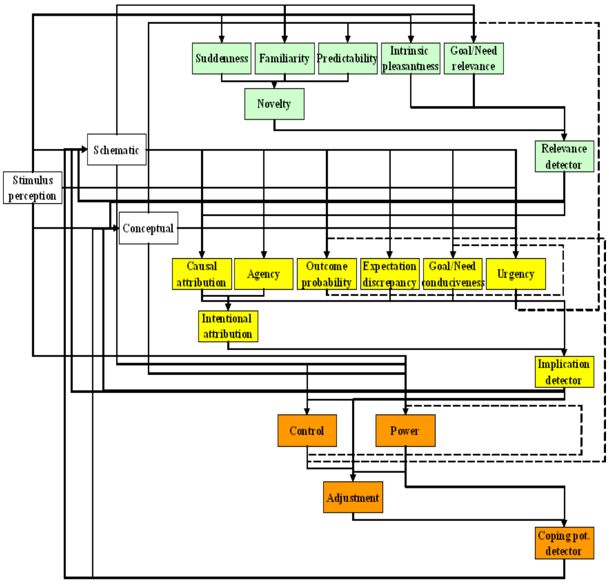


Figure 3. Graphical representation of the SSK model.

We based our integration on two common architectural concepts of the models: first, the separation of appraisal into three distinct levels of information processing (represented in Fig 3 by (a) the three white perception processes that influence processing in different appraisal processes, and (b) the layers of appraisal processes that trigger each other predominantly from stimulus oriented to conceptual), and second, the appraisal registers/detectors (the three appraisal integrators (mediating processes) at the right of Fig. 3). In our integration we chose to focus on processes and their dependencies. In [Broekens et al., *submitted*] we show in more detail that the common formal notation briefly introduced above can be used as a tool to compare and integrate different appraisal theories. Both are important reasons to formalize appraisal theories. Integration was greatly facilitated by the formalism's ability to describe in detail the processes, their conditional dependencies based on second-order predicates and the appraisal-dimensions.

## 5 Facilitation of computational model evaluation

Our formalism helped to develop a computational model based on the newly created 'theory of appraisal', the SSK model [Broekens et al., *submitted*]. It facilitated (1) filling in computational details, and

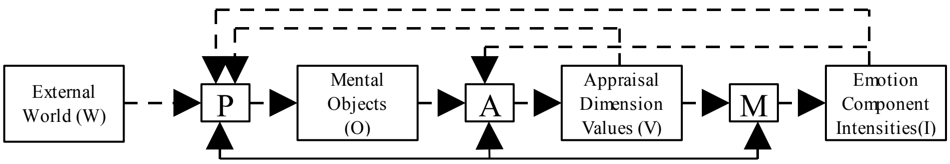


Figure 2. Formalism overview. Dotted arrows denote process input, normal arrows are potential process dependencies.

(2) making computational assumptions explicit. Further, the formal description helped us to evaluate our computational model with respect to the SSK model. This research shows that a common formalism for the structure of appraisal is useful for the specification and evaluation of a computational model.

## 6 Summary and conclusion

Analysis of the relation between structural, process and computational models of appraisal shows that formalization of structural models is needed.

*First, to advance appraisal theory*. A formal description facilitates comparison, convergence and integration of appraisal theories, and the process of formalization helps theory refinement.

*Second, to develop and evaluate computational models of emotion based on structural theories of appraisal*. First, we argued that process models of appraisals should coexist with computational models (CM), not take their place. Second, before designing CMs at the algorithmic level, declarative information is needed on the processes that are responsible for perception, appraisal and mediation as defined by the appraisal theory. Third, objective information is needed to evaluate the consistency between CM and appraisal theory, and reuse of experimental findings seems essential; a declarative description of the processes that are responsible for an agent's emotion is needed to check if the agent's unexpected emotion resulting from an experimental situation is due to a problem in the agent's architecture, or due to a mismatch between our interpretation of the situation and the agent's.

Our research shows that a common formalism facilitates integration between appraisal theories, and is a tool to narrow the gap between structural models of appraisal and computational models.

## Acknowledgments

The authors wish to express their gratitude to Drs Walter Kosters and Fons Verbeek for their helpful suggestions and critical reading.

## References

[Baars, B. J., & Franklin, S., 2003]. How conscious experience and working memory interact. *TICS* 7, (4), 166-172.

[Breazeal, C., 2001]. Affective Interaction between Humans and Robots. In J. Keleman & P. Sosik (Eds), *ECAL 2001, LNAI 2159* (pp. 582-591).

[Broekens, J. et al., to appear]. Formal Models of Emotion: Theory, Specification, and Computational Model. *(submitted, draft version upon request).*

[Broekens, J., & DeGroot, D., 2004]. Emotional Agents need Formal Models of Emotion. In R. Verbrugge, N. Taatgen & L. Schomaker (Eds.), *16th Belgian-Dutch Conference on Artificial Intelligence (BNAIC)*, (pp.195-202).

[Coddington, A., & Luck, M. 2003]. Towards motivation-based plan evaluation. In I. Russell, & S. Haller, S., (Eds.). *Proc. of the Sixteenth International FLAIRS Conference*, (pp. 298-302).

[Franklin, S., & Graesser, A., 1999]. A software agent model of consciousness. *Consciousness and Cognition*, 8(3), 285-301.

[Frijda, N. H., & Mesquita, B., 2000]. Beliefs through emotions. In: N. H. Frijda, A. S. R. Manstead & S. Bem (Eds.). *Emotions and Beliefs*.

[Gratch, J., & Marsella, S., 2004]. A domain independent framework for modeling emotion. Journal of *Cognitive Systems Research*, 5 (4), 269-306.

[Henninger, A.E., Jones R.M., & Chown, E., 2003]. Behaviors that emerge from emotion and cognition: implementation and evaluation of a symbolic-connectionist architecture. *AAMAS 2003* (pp. 321-328).

[Heylen, D., Nijholt, A., Akker op den, R., & Vissers, M., 2003]. Socially intelligent tutor agents. *IVA 2003,* (pp. 341-347).

[IEEE, 2000], *IEEE Recommended Practice for Architectural Description of Software-Intensive Systems, Std 1471-2000*, IEEE.

[Mallery, J. C., 1988]. *Thinking about Foreign Policy: Finding an Appropriate Role for Artificially Intelligent Computers.* Cambridge: Master's Thesis, M.I.T. Political Science Department.

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

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

[Reisenzein, R., 2001]. 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*.

[Roseman, I. J., & Smith, C. A., 2001]. Appraisal theories: Overview, assumptions, varieties, controversies. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.). *Appraisal Processes in Emotion: Theory, Methods, Research*.

[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*.

[Smith, C. A., & Kirby, L. D., 2000]. Consequences require antecedents: Toward a process model of emotion elicitation. In J. P. Forgas (Ed.). *Feeling and Thinking: The role of Affect in Social Cognition.*

[Wehrle, T., & Scherer, K. R., 2001]. Towards computational modeling of appraisal theories. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.). *Appraisal Processes in Emotion: Theory, Methods, Research*.