Computational Modeling of Emotion: Towards Improving the Inter- and Intradisciplinary Exchange

Rainer Reisenzein, Eva Hudlicka, Mehdi Dastani, Jonathan Gratch, Koen Hindriks, Emiliano Lorini, John-Jules Meyer

Computational Modeling of Emotion: Toward Improving the Inter- and Intradisciplinary Exchange

Rainer Reisenzein, Eva Hudlicka, Mehdi Dastani, Jonathan Gratch, *Member*, *IEEE*, Koen Hindriks, Emiliano Lorini, and John-Jules Ch. Meyer

Abstract—The past years have seen increasing cooperation between psychology and computer science in the field of computational modeling of emotion. However, to realize its potential, the exchange between the two disciplines, as well as the intradisciplinary coordination, should be further improved. We make three proposals for how this could be achieved. The proposals refer to: 1) systematizing and classifying the assumptions of psychological emotion theories; 2) formalizing emotion theories in implementation-independent formal languages (set theory, agent logics); and 3) modeling emotions using general cognitive architectures (such as Soar and ACT-R), general agent architectures (such as the BDI architecture) or general-purpose affective agent architectures. These proposals share two overarching themes. The first is a proposal for modularization: deconstruct emotion theories into basic assumptions; modularize architectures. The second is a proposal for unification and standardization: Translate different emotion theories into a common informal conceptual system or a formal language, or implement them in a common architecture.

Index Terms—Computational emotion modeling, affective computing, emotion theories, theory formalization, agent logics, cognitive architectures

1 Computational Modeling of Emotion as an Interdisciplinary Project

1.1 Definition and Aims of Computational Modeling of Emotion

As used in this paper, *computational modeling of emotion* refers to attempts to develop and validate computational models of human emotion mechanisms. Computational modeling of emotion is an interdisciplinary endeavor between, in particular, psychology and computer science [135]. The goals of computational modeling of emotion largely correspond to the general goals of AI (e.g., [134]), when these are restricted to the domain of emotions: To achieve a better theoretical understanding of emotions in natural (specifically human) and artificial agents by creating computational models of them; and to enrich the architecture of artificial agents with emotion mechanisms similar to those of humans, and thus with the capacity to "have" emotions.[1] Computational emotion modeling projects that have the first goal can be called *theoretical* and those that have the second goal *applied* [31]. Psychologists are typically more interested in theoretical and computer scientists in applied computational emotion modeling; however, some researchers in both disciplines pursue both goals. In fact, the two goals are closely linked: Endowing artificial agents with truly human-like emotion mechanisms presupposes reasonably faithful computational models of these mechanisms (e.g., [68]); conversely, one (some would say the best) way to attain a deep theoretical understanding of mental processes, including emotions, is to try to synthesize them in artificial agents (e.g., [25]).[2]

- *R. Reisenzein is with the Institute of Psychology, University of Greifswald, Franz-Mehring-Str. 47, 17487 Greifswald, Germany. E-mail: rainer.reisenzein@uni-greifswald.de.*
- *E. Hudlicka is with Psychometrix Associates Inc., PO Box 9409, North Amherst, MA 01059. E-mail: hudlicka@ieee.org.*
- *M. Dastani and J.-J. Ch. Meyer are with the Department of Information and Computing Sciences, University of Utrecht, PO Box 80.089, TB Utrecht 33507, The Netherlands. J.-J. Ch. Meyer is also with the Alan Turing Institute Almere. E-mail: {M.M.Dastani, J.J.C.Meyer}@uu.nl.*
- *J. Gratch is with the Institute of Creative Technologies, University of Southern California, 12015 Waterfront Way, Playa Vista, CA 9009. E-mail: gratch@itc.usc.edu.*
- *K. Hindriks is with the Department of Intelligent Systems, Delft University of Technology, PO Box 5031, 2600 GA Delft, The Netherlands. E-mail: K.V.Hindriks@tudelft.nl.*
- *E. Lorini is with the Centre National de la Recherche Scientifique Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, Cedex 9, France. E-mail: Emiliano.Lorini@irit.fr.*

1. Based on a functionalist view of mental states (e.g., [24]), we propose to understand this capacity as the capacity of artificial agents to have internal states that are functionally equivalent or at least similar to emotions in humans, i.e., that play causal roles in the agent architecture that mimic those played by emotions in humans. The causal effects of emotions that can be modeled in artificial agents include self-awareness of emotions when they occur (e.g., [20], [135]); however, no claim is made that artificial agents, at least those that currently exist, are conscious of their emotions in the sense of having qualitative phenomenal experiences (see [154]).

2. An additional goal of affective computing is to provide artificial agents with a theory of mind module that allows them to recognize, explain, predict, and conceive of ways to influence, the emotions of other agents ([135]; see also [152]). Computational models of emotion can be helpful for this purpose, too: Declarative versions of these models (possibly in a simplified form) can be used as components of the theory-of-mind module of artificial agents; alternatively or in addition, agents can be programmed to use their emotion mechanisms in "simulation mode" to predict the emotions of others (as well as their own) (as realized, for example, in Pynadath and Marsella's [139] *PsychSim* agent).

The close connection between theoretical and applied computational emotion modeling is one reason why the past years have seen increasing cooperation between psychology and computer science in research on emotions (see, e.g., [11], [78], [168]). Some see in this development a revival of the cognitive revolution of the 1960s (e.g., [116], [119]) in the cloak of affect [67]. While this may seem to be a grand claim, there is at least one sense in which it seems justified: Because emotions are thought to be primarily useful for resource-bounded autonomous agents acting in dynamic and only partially known environments (e.g., [118], [135], [100], [179]), the computational modeling of emotion requires returning to the original goal of artificial intelligence to create "complete" artificial agents (see, e.g., [89], [164]); and likewise, a refocusing on the original goal of cognitive science, to create a unified theory of the mind [121]. The cooperation between psychology and computer science in the field of emotion therefore has the potential to contribute to unification, not only in emotion psychology, but in psychology at large.

1.2 Realizing the Potential of Computational Emotion Modeling

However, to realize its potential, the exchange between as well as the coordination within the two disciplines should be further improved. As to the *interdisciplinary* exchange, the export of theories has so far been mainly from psychology to affective computing, rarely in the reverse direction (see [31]). For example, as Broekens [31] notes, the *Handbook of Emotions* [94], a standard psychological reference work on emotions, contains virtually no citations to the affective computing literature. This is so despite the fact 1) that psychologists frequently emphasize the need to better understand the computational mechanisms underlying emotions (e.g., [167]) and 2) that at least some of the emotion models developed by computer scientists can claim to be serious candidates for computational theories of human emotion, or at least to elucidate and concretize the psychological theories on which they are based (for supportive evidence see Sections 3 and 4, and [31], [108]). Concerning the *intradisciplinary* situation, both disciplines could profit from more comparative efforts, coordination, systematization, and focus. Within affective computing research, there are now numerous computational models of emotions (for overviews, see, e.g., [78], [108]), but this work is rarely cumulative or even comparative; rather, as noted by Marsella et al. [108], different researchers have tended to start anew from the same or different psychological sources (for notable exceptions, see [27], [53]). Partly as a consequence, the theoretical landscape in affective computing increasingly resembles that in the psychology of emotion, where numerous psychological theories of emotion coexist side by side (see, e.g., [155], [167]), with few attempts to integrate or even to systematically compare them (for an exception, see [148]). From the psychologist's perspective, affective computing researchers may even seem to further contribute to theoretical fragmentation by developing different computational models of the same psychological theory (as documented, for example, by the many implementations of the OCC theory of emotion [129]).

1.3 Aims and Overview of This Paper

Our aim in this paper is to make several proposals for further improving the interdisciplinary exchange between, as well as the intradisciplinary coordination within, psychology and artificial intelligence research in the domain of emotion modeling. A complementary perspective—in the form of a set of guidelines for prospective computational emotion modelers—is provided by Hudlicka [80]. In Section 2, we make proposals for systematizing the psychological theories of emotion that typically serve as blueprints for creating computational emotion models. These proposals can also be helpful for systematizing and comparing existing computational emotion models. In Section 3, we discuss possibilities of formalizing psychological (and philosophical) emotion theories in implementation-independent formal languages. Again, these languages can also be useful for formalizing (at least part of) the assumptions of existing computational models. In Section 4, we consider options for modeling emotions in existing, comparatively general-purpose cognitive and agent architectures, including general-purpose affective agent architectures. Due mainly to the composition of the group of authors, our discussion focuses on *symbolic* models of *psychological* (specifically cognitive) emotion theories.[3] Hence, we ignore the computational modeling of neurophysiological emotion theories (e.g., [8]) as well as attempts to implement psychological emotion theories in subsymbolic neural networks (e.g., [12]). Note, however, that the symbolic modeling of psychological emotion theories is by far the dominant approach in the field of computational modeling of emotion [78], [108].

2 Systematizing Psychological Emotion Theories

2.1 The Situation in Emotion Psychology

From the perspective of an applied affective computing researcher, psychology would ideally be at an advanced stage where *the ultimate correct theory of emotion* (UCTE) or even better, the *ultimate unified theory of the mind* (UUTM) had been attained. Furthermore, that theory would ideally be formulated as a computational model, or would at least be available in a format that lends itself readily to implementation as a computer program. Unfortunately, psychology has not yet arrived at this stage. First, there is not one emotion theory; there are many theories: Strongman [190] lists no less than 150 psychological and philosophical emotion theories that have been proposed during history, and these are still not all. To be sure, some of these "theories" are no more than vague hypotheses, there is much redundancy, and many theories are of historical interest only. However, even if one restricts attention to the currently discussed emotion theories, many do remain (e.g., [80], [155], [167]). Making matters worse, the exact relations between the different emotion theories are sometimes difficult to judge even by experts. Second, the existing

3. Although the views expressed in this paper mostly reflect a majority consensus, agreement among all authors on all points is not implied. However, considering the many issues discussed in this paper and the different backgrounds and theoretical predilections of the authors, it was encouraging to see how much agreement could be achieved.

psychological emotion theories are not formulated as computational models and most are not described in a sufficiently precise and complete way to be directly implemented as computer programs.[4]

Theoretical fragmentation and a comparative lack of precision are not peculiar to emotion psychology but are by and large characteristic of psychology in general. Psychologists often excuse these problems of their field by pointing to the complexity of psychology's subject matter and the comparatively young age of psychology as a science. There is truth to these claims. However, there are also other reasons for the present state of psychology: Partly as an aftereffect of the behavioristic phase in psychology (about 1920-1960) with its associated positivist philosophy of science, a premium is still placed in today's psychology on empirical as compared to theoretical research [177].[5] Furthermore, novelty and surprisingness of claims and findings are honored more than are attempts to integrate and unify. The recent interaction between psychology and computer science in the field of emotion has created a need to provide concise and precise summaries of psychological theories to nonspecialists who are at the same time accustomed to a high degree of precision. This need could provide an impetus to psychologists to try to clean up and organize their theoretical landscape. In this, they can be helped by computer scientists and logicians (see [41], [80], and Section 3).

2.2 Toward Systematization and Unification in Emotion Psychology

There are several, not necessarily exclusive ways how this aim could be achieved. One possibility is to systematically reconstruct the existing emotion theories; possibly using a formal language (e.g., [204]; see Section 3 for more details). However, this approach, although appropriate for selected theories of particular importance, is probably too cumbersome as a general strategy. Not every theory needs to be precisely reconstructed; in particular, it is not necessary to formalize theories that have turned out to be (or at least, most likely are) false. One should focus on the promising alternatives.

An alternative, quicker and more viable strategy may be to 1) break up existing emotion theories into their component assumptions and 2) reformulate these assumptions in a common conceptual framework (e.g., [80], [108], [147]). This common framework could be a formal language, possibly enriched with a set of special concepts (e.g., *belief* and *desire*) or it could be a general-purpose cognitive architecture (see Sections 3 and 4); however, already semiformal or informal theory reconstructions can significantly contribute to clarification and unification if concepts are used in a clear and consistent way [41]. The ultimate aim of this second strategy would be to reduce the many existing emotion theories to a smaller (hopefully *much* smaller) set of basic assumptions. These building blocks of emotion theories can then be separately discussed and refined. Furthermore, similar assumptions of different emotion theories can be more easily compared, vague assumptions can be more easily clarified, and conflicting assumptions more easily identified, when they are recast in a common conceptual framework (see also, Section 3 and 4).

In the remainder of Sections 2, we make proposals for breaking up and organizing the assumptions of emotion theories in a principled way. These proposals derive from a reflection on the basic task of emotion psychology. The mainstream position in today's psychology, at least its cognitive branch—which is not only compatible with, but in fact rests on, a computational view of the mind (see footnote 4)—conceives of humans as autonomous cognitive systems, or intelligent agents, whose basic architecture emerged in evolution, but is significantly elaborated and possibly even transformed, by individual and cultural learning. This theoretical perspective implies that, to explain human behavior and experience, it is necessary to decode the structure and functioning of the human information processing system. For emotion psychology, this means that its central task is to reverse-engineer the structure and functioning of the human emotion system, including its connections to other subsystems of the mind (e.g., [155]; as mentioned in Section 1.1, to understand emotions, it may in fact be necessary to reconstruct significant parts of the whole cognitive architecture). As already emphasized by McDougall [110], the task of reverse-engineering the emotion system comprises identifying the inherited versus acquired components of the emotion system.

2.2.1 On the Definition of Emotion

Before proceeding, a note on the concept of emotion is in order. Just as there is at present no generally accepted *theory* of emotion, there is currently no generally accepted *definition* of emotion—at least if by that one means a specification of the singularly necessary and jointly sufficient features of emotion. Causal-role functionalism about mental states ([24], footnote 1) suggests that the latter is a direct consequence of the former [149], [157]. According to functionalism, mental states are defined (at least in part) in terms of their causal-functional role in the mental economy. Because this role is specified by a theory of emotion, a precise definition of emotion can only be given in the context of an emotion theory that is presupposed as correct. Hence, a precise definition of emotion is necessarily a *theoretical* (theory-based) *definition*. In addition, because a theoretical definition of emotion stands or falls with the emotion theory that it presupposes, it is not a linguistic convention or stipulation about the use of the term "emotion," but an empirical hypothesis about the nature of emotions [149], [157].

However, a precise definition of emotion is not needed to conduct meaningful research on emotions. All that is needed for this purpose is a rough, preliminary demarcation of the object domain "emotion"—what can be called a

4. This is actually surprising if one considers that, since the cognitive revolution of the 1960s, the psychological mainstream (at least in cognitive psychology) subscribes officially to the "information processing paradigm", that takes information processing in computers as the guiding metaphor for understanding mental processes in humans. According to the information-processing paradigm, the task of (cognitive) psychology is to reconstruct the "rules and representations" that underlie mental processes [131]. In practice, however, the development of theories of mental processes that fully meet this description (i.e., computational models) in psychology is still restricted to a comparatively small community of experts (see, e.g., [191] for an overview) and only very few of them are working on computational models of emotion.

5. One indicator of this "bias for the empirical" is that, different from most other sciences, psychology has no official theoretical branch: There is no official subdiscipline called *Theoretical Psychology* [177].

working definition of emotion [149], [157]. Minimally, such a working definition is already provided by a list of typical examples of emotion. We believe that, despite their differences of opinion regarding the true nature or correct theoretical definition of emotions, most emotion researchers are by and large in agreement on such a list. That is, most emotion researchers would agree among themselves, as well as with lay persons, that the objects of their inquiry comprise, centrally, those mental (or perhaps mental-behavioral) states of humans that in common language are denoted by words such as "joy," "sadness," "fear," "anger," "pity," "pride," and so on. Beyond this, a working definition of emotion can include those features of emotion that are largely uncontroversial and, therefore, do not imply a decision for or against a particular emotion theory. Although there is perhaps no feature of emotion that has not been questioned by one or the other theorist, it is widely accepted 1) that emotional states (joy, fear, etc.) are normally reactions to the perception, imagination or thought of certain things (typically events or states of affairs); and 2) that emotions have characteristic subjective as well as objective aspects. Subjectively, emotions manifest themselves, centrally, in characteristic phenomenal experiences or feelings, that appear to be directed at their eliciting objects: For example, one feels happy about the arrival of a friend, or experiences surprise about the election victory of a political party. Objectively, emotions manifest themselves at least occasionally in certain characteristic actions (e.g., flight or avoidance in the case of fear), expressive reactions (e.g., smiling in joy), and physiological changes (e.g., a rise of blood pressure in anger).

Although the proposed working definition of emotion is necessarily fuzzy, this is no objection to it. Just as providing a precise definition of emotions is the result rather than the precondition of theorizing about emotions, so too is providing a sharp demarcation of the domain of emotions, as well as a principled taxonomy of emotions (see also [129], [150]).

In this paper, we will, however, make one additional assumption about emotions: We will assume that whatever emotions are more precisely, they are *mental states* which are caused by, and in turn cause, other mental states and behaviors. Harmless and even self-evident as this assumption may seem to some readers, it is (at first sight at least) in conflict with those emotion theories that view emotions as *response syndromes* including both mental and behavioral components (e.g., [92], [166]; see [149] for a discussion). However, the syndrome theories, too, assume a generating process that maps (representations of) objects or events into one or several emotion-specific mental states (e.g., appraisals, bodily feelings, action tendencies), which then, singularly or in combination, influence further cognitive processes and behavior. With respect to the syndrome theories, our term "emotion" refers to the emotion-specific internal states assumed in these theories.

2.2.2 Reverse-Engineering the Emotion System: "Horizontal" and "Vertical" Task Divisions

The task of emotion psychology—the reverse-engineering of the emotion system—can be broken down both "horizontally" and "vertically": Horizontally, by following the temporal and causal flow of information from the perception (or an internally generated representation) of an eliciting object to emotion, and from there to the cognitive and behavioral effects of emotion; vertically, by studying the emotion system at different levels of analysis. In particular, one can focus on either the "intentional level" of system analysis or on the "design level" (Dennett, [51]; similar distinctions have been drawn by Marr [106] between the computational and algorithmic level, and by Newell [122] between the knowledge level and the symbol level).

The proposed *horizontal division* suggests three main questions for emotion psychology: 1) How are emotions generated? 2) What are their effects on subsequent cognitive processes and behaviors? and 3) related to both 1 and 2 (see below), what is the nature of the emotion itself? ([80], [108], [155], see Section 2.2.1).

The *vertical division*, between the intentional and the design level of system analysis, separates 1) theoretical descriptions of emotion processes in terms of what are ultimately common-sense psychological concepts (*perception, belief, desire, intention, fear, hope,* etc.) from 2) theories that (also) refer to the internal representations (sentences in a language of thought, images, subsymbolic signals, etc.) and computations on these representations, that according to the computational view of the mind underlie the intentional-level phenomena [51]. The term "intentional" in Dennett's distinction expresses the idea that intentional-level theories explain the behavior of a system by attributing to it *intentional* (meaning here: *representational*) *mental states* familiar from common-sense psychology, such as beliefs and desires (e.g., [70], [171]): Internal system states that refer to or represent aspects of the world, including the own system, in particular ways (e.g., as being true in the case of belief, or as to-be-made true if possible in the case of desire). In adopting the "intentional stance" toward a system [51], we assume—either seriously or only as an "as-if" explanatory strategy—that the behavior of the system is controlled by representational mental states of this kind [70], [171].[6]

Simple as it is, the proposed classification scheme brings several important facts about psychological emotion theories into focus. Concentrating on the vertical division between the intentional and the design level of system analysis brings out that, as already mentioned, the majority of the existing psychological emotion theories are formulated on the intentional level (e.g., "fear is experienced if the person believes that an undesired event might occur"). Nevertheless, it should be noted that some psychological emotion theories (e.g., [62], [151], [166], [181]) provide at least sketches of the computational processes assumed to underlie the intentional-level phenomena. That is, whereas "pure" intentional-level theories of emotion only describe inputs and outputs of presumed mental computations, these "process theories" also make some assumptions about the computational processes that transform inputs into outputs (e.g., the nature of the underlying representations,

6. Note that the distinction between the intentional and the design level is orthogonal to the distinction between deliberate/conscious and automatic/unconscious mental processes referred to later in our discussion of cognitive versus noncognitive emotion elicitation. Both deliberate/conscious and automatic/unconscious emotion generation processes can in principle involve folk-psychological mental states such as beliefs and desires (e.g., [33]); and both processes can also be described in terms of representations and computations.

the identity and sequence of the involved subprocesses; see [147]). As such, these process theories can play the role of intermediaries between intentional-level theories of emotion and full-fledged computational theories [32], [100].

Other important facts about emotion theories become transparent if one focuses on the proposed horizontal division, the flow of processing. First, quite a few historically important and contemporary emotion theories are primarily concerned with the question of the *nature* of emotion and say comparatively little (at least little specific) about how the different emotions are produced, and what useful or other effects they have (see [154]). These theories include James's [82] theory of emotion and neo-Jamesian emotion theories (e.g., [86]); cognition-arousal theory [16], [101], [165] and "dimensional" emotion theories such as the pleasure-arousal model [144], [162], [163]. The question that these theories mainly seek to answer is what kind of mental states emotions are, which is interpreted as meaning: From which building blocks, if any, emotions are constructed, and how they fit into a general taxonomy of mental states. Specifically, are emotions mental states *sui generis* (e.g., centrally generated feelings of pleasure and displeasure) or can they be reduced, at least in part, to known mental states (e.g., feelings of bodily changes, a combination of cognitions and bodily feelings, patterns of appraisal or of beliefs and desires, or felt action tendencies; see [154])? Theorists interested in these questions have typically proceeded on the assumption that they can be answered without *detailed* theories of the generation and the effects of emotion (although *some* causal assumptions seem unavoidable). However, because of their restricted focus, the resulting theories of the nature of emotion are alone insufficient as psychological foundations of computational emotion models; they must be combined with theories of the elicitation and the effects of emotions. This becomes evident, at the latest, when one tries to use these theories to construct emotion mechanisms for artificial agents. For example, Becker-Asano and Wachsmuth [20] have implemented a version of the pleasure-arousal theory of emotion in an artificial agent; however, to account for the generation of emotions, they found it necessary to enrich pleasure-arousal theory with the OCC theory [129].

Second, even theories that are quite specific about emotion generation (in particular, some cognitive emotion theories) are sometimes relatively silent about the *effects* of emotions (see [80]). Again, if these theories are to be used as starting points for creating emotion mechanisms for artificial agents—which are usually desired because one hopes that they will enhance the functioning of the agents in some way—they must be enriched with assumptions about the effects of emotions (see, e.g., [20], [68], [53]).

Interestingly, most theories of emotion generation seem to be compatible with rather different assumptions about the nature of emotions and their effects. For example, different cognitive emotion theorists have proposed that emotions are bodily feelings, feelings of pleasure-displeasure, mixtures of pleasure-displeasure and activation-deactivation, felt action tendencies, evaluative beliefs, patterns of belief and desire, various combinations of these states, and even syndromes including both mental and bodily reactions (see Section 2.1.1). Likewise, most theories of emotion generation seem to be compatible, in principle, with a variety of hypotheses about how emotions influence subsequent cognitive processes and actions: by creating hedonistic desires, by creating nonhedonistic action tendencies, by focusing attention or shifting goal priorities, and so on (see Section 2.2.4). Furthermore, most cognitive theories of emotion generation seem equally compatible with a highly modular, "discrete emotions" view of the emotion system (e.g., [92], [161]) and with a fairly domain-general view (e.g., [150], [151], [163]). The compatibility of the assumptions of different emotion theories about the generation, nature, and effects of emotions may seem desirable from an engineering perspective, but it can be argued that it actually reveals a weakness of the respective emotion theories: Apparently, they are unable to impose strong constraints on what goes with what.[7]

2.2.3 Classifying Theories of Emotion Elicitation

Cognitive versus noncognitive theories of emotion. The assumptions of the different existing psychological (and philosophical) emotion theories about the process of emotion elicitation can be classified into two main groups, "cognitive" and "noncognitive." Although the usefulness of this distinction has been questioned because of the vagueness and ambiguity of the term "cognitive," it seems that a meaningful dividing line can still be drawn as follows: Cognitive theories assume that emotions (or at least a core subset of emotions) require certain "higher-order" mental representations, in particular, they require what are commonsensically called *beliefs* and *desires* [70], [111].[8] In their more extreme forms, cognitive theories of emotion even assume that emotions *are* a form of cognition (e.g., evaluative beliefs, or belief-desire compounds), or contain beliefs or belief-desire compounds as essential components (see [37], [153], [154]). In contrast, "noncognitive" emotion theories assume that 1) certain kinds of emotions in a broad sense (e.g., sensory affects such as the hedonic pleasure evoked by the smell of a rose), or even 2) some prototypical emotions such as fear, are always (sensory affects) or at least sometimes (fear) elicited by stimuli via a more direct route, that circumvents the mentioned higher cognitive processes. According to the most interesting interpretation of this hypothesis (e.g., [82], [128]), this "noncognitive" pathway involves only *nonpropositional* and possibly even *nonconceptual* mental representations, such as images, or simple sensations.

Noncognitive emotion elicitation in computational models. There is a longstanding and still ongoing discussion in psychology about the importance, and even existence, of "noncognitive" emotion generation (e.g., [91], [93], [189], [209]). This discussion has suffered, among other things,

7. Closer investigation of the issue suggests, however, that there are at least tensions between the assumptions of different emotion theories about the causes, nature and effects of emotions. For example, it can be argued that, at closer look, a cognitive view of emotion elicitation does not fit comfortably with a discrete emotions perspective [148].

8. Alternative terms used for "desire" in the emotion literature are "motive" (e.g., [92]), "goal" (e.g., [129], [166]) and "concern" (e.g., [62]). Although all of these terms are meant to denote similar *motivational representations*, there are some differences between theorists about how, exactly, they are conceptualized. Here, we understand desires as motivational representations with propositional contents, i.e., as motivational representations of states of affairs [70], [171]. Desires include both performative desires (desires to do something) and nonperformative desires (desires that something should be the case) [195].

from the failure to distinguish clearly between the two described version of the hypothesis of noncognitive emotion elicitation. Hypothesis (a), which claims that certain kinds of affect (such as sensory pleasures and displeasures) are noncognitively generated in the described sense—i.e., presuppose only nonpropositional or even just nonconceptual representations—is intuitively plausible [148]. In contrast, Hypothesis (b), according to which even prototypical emotions such as fear, anger, or joy can be (and perhaps even, often are) noncognitively caused (e.g., that fear can be elicited by the visual appearance of an abyss or by loud sounds, without any mediating thoughts), is more controversial. On closer inspection, the data that have been adduced to support this hypothesis turn out to be less convincing than is often claimed (see, e.g., [151], [175], [173]). For example, it has been argued that noncognitive emotion generation is demonstrated by the finding that certain emotional stimuli elicit physiological reactions even if they are presented below the threshold of awareness (e.g., [128]; see [189] for a review). However, recent studies point to the possibility that the observed physiological reactions were shown by a subset of the participants who were still able to recognize the emotional stimuli at the presumed "subliminal" exposure durations (see, e.g., [132]). Furthermore, even if it is a genuine phenomenon, emotion elicitation by subliminal stimuli could be mediated by automatized and unconscious cognitive appraisal processes (e.g., [174], [181]).

It should be noted, however, that regardless of how this issue is decided, the computational modeling of "reactive" emotion elicitation by specific objects or patterns of sensory stimulation is a comparatively easy task (e.g., [3], [53]). The real computational challenge is posed by the cognitively mediated emotions [68].

Cognitive emotion elicitation: two theories. Philosophers of emotion (e.g., [70]) have drawn a distinction between two basic kinds of cognitive emotion theories: *cognitive-evaluative* (or evaluative-beliefs) theories, and *cognitive-motivational* (or belief-desire) theories. The crucial difference between the two concerns the proximate mental causes (or, depending on viewpoint, the constituents; [154], [153]) of emotions: Are they only beliefs (including evaluative beliefs); or are they beliefs and desires? To illustrate the difference with a simple example, according to the cognitive-evaluative theory (which arguably comprises most psychological appraisal theories), John is happy about a state of affairs p if he believes that p is the case and believes that p is good for him (e.g., [9]). In contrast, according to the belief-desire theory, John is happy about p if he believes p and desires p (e.g., [37], [70], [126], [151]). Hence, according to the belief-desire theory, the evaluative beliefs (appraisals) featured in the cognitive-evaluative theory of emotion are in fact not required for emotions; all that is needed are factual beliefs and desires.

The distinction between the two theories rests on the assumption, which is widely shared in contemporary philosophy, that beliefs and desires are fundamentally distinct kinds of mental states that cannot be reduced to one another (e.g., [70], [171]). Beliefs are *informational* mental states: their function is to provide the agent with information about the world and the self. By contrast, desires are *motivational* states: their function is to provide the agent with goals, states of affairs whose realization is attempted under suitable circumstances (e.g., [70]; see also [143]).

Proponents of the belief-desire theory of emotion have argued that this theory avoids a number of objections to the cognitive-evaluative theory (see [70], [151]), including the objection that evaluative beliefs are, as a matter of fact, neither necessary nor sufficient for emotions. A discussion of these arguments, which would require an in-depth comparison of the foundational assumptions of the two kinds of cognitive emotion theory, is beyond the scope of this paper. However, we would like to point out that it is possible to a large degree to "map" the assumptions of each theory into the conceptual framework provided by the other (although with a corresponding shift in meaning!). This remapping is possible because most appraisal theories assume that appraisals (evaluations of perceived events as good or bad) result from the comparison of perceived or believed events with goals [148]. Psychological appraisal theory can, therefore, be regarded as a special version of the belief-desire theory of emotion which assumes that the causal link between beliefs and desires on the one hand (e.g., the belief that p and the desire for p), and emotions on the other hand (e.g., the emotion of joy about p) is mediated by the appraisal of p [148]. Conversely, the belief-desire theory can be recovered from appraisal theory by linking belief and desire directly to the emotion. This can be achieved by either *dropping* the mediating appraisal from the theory, or by *reinterpreting* the appraisal (the outcome of the comparisons of perceived events with goals) as *being the emotion* (e.g., as a preverbal form of evaluation that is experienced as a feeling of pleasure or displeasure; see [151]). Several recent formalizations of appraisal theories have in fact implicitly used this approach and thus have effectively reinterpreted appraisal theory as belief-desire theory (e.g., [1], [187]; see Section 3.2). Analogously, some computational models of emotion (e.g., [20]) treat emotions as direct effects of goals and beliefs, without a separate, mediating appraisal state (see Section 4.2).

2.2.4 Classifying Hypotheses about Emotion Effects

Compared to the broad consensus about how emotions are elicited, there is considerably less agreement in the psychological (and philosophical) literature about how, exactly, emotions influence subsequent cognitive processes and behavior. Many different hypotheses about the effects of emotions have been proposed (see, e.g., [80] for an overview). To get a systematic grip on them, it is useful to distinguish, first, between *functional* and *nonfunctional* effects of emotions. The concept "functional" in this distinction is best explicated in evolutionary terms (e.g., [117]): Functional effects of emotions are those of their effects that were responsible for the selection of the corresponding emotion mechanisms in evolution, whereas nonfunctional effects of emotions are neutral, or even negative effects of emotions that are tolerated to reap the benefits of emotions (see also [120]).[9] To illustrate, one proposed functional effect of emotions is the focusing of attention on the emotion-evoking events (see below). A nonfunctional and potentially even harmful side effect of this is a reduced capacity for engaging in other, simultaneous

9. The assumption that emotions are overall adaptive is widely accepted in today's psychology. The best general argument for this assumption is that it is hard to see how complex mechanisms like emotions could have emerged in evolution other than by natural selection, and hence because of their useful effects (see [49]).

tasks. Ideally, only functional effects of emotions would have to be *explicitly* considered in computational emotion models, because nonfunctional effects would arise naturally as side effects of the activity of the mechanisms that produce the functional effects (e.g., [160]).

Although many specific functions of emotions have been proposed (e.g., [63], [70], [85]), it appears that most of them can be classified into one of three categories (e.g., [130]):[10]

1. *Informational or epistemic functions of emotions.* According to this hypothesis, a main function of emotions is to provide adaptively useful information to other cognitive (sub-)systems, including other agents (via nonverbal or verbal communication). The information in question is held to be in particular information about a) the occurrence of appraisals (e.g., [170]) or of changes in the person's belief-desire system [151], [157] and b) related to this, information about the value of objects and events, including action consequences (e.g., [43], [111], [180]).
2. *Attentional or resource-allocation functions of emotions*. According to this hypothesis, emotions serve to shift the focus of attention to (or computationally speaking, allocate central processing resources to) their eliciting events, or to give priority to them in processing (e.g., [176], [151], [178]).
3. *Motivational functions of emotions*. This hypothesis claims that a main adaptive function of emotion is to reprioritize existing or to generate new goals or intentions (e.g., [62], [127]). With respect to the generation of new goals by emotions, the two main proposals are a) emotions create hedonistic desires (desires to reduce or avoid negative, or to maintain or produce positive feelings; e.g., [18], [113]); b) emotions induce emotion-specific goals or action tendencies (e.g., fear causes the desire to flee, anger to aggress, pity to help) *directly*, that is, without the mediation of hedonistic desires (e.g., [62], [109], [161], [203]). A possible mediating mechanism for the latter effect could be that emotions draw attention to, or increase the salience or persistence of goals currently at stake (e.g., [3], [77], [118]; see also [145]). Hence, the motivational effects of emotions could be partly based on their attentional and informational effects [151].

2.2.5 A Note on the Relation between Emotion and Motivation

We end this part with a few comments on the relation between the psychology of emotion and motivation (including decision making).[11] Although there is general agreement that emotion and motivation are related and several early theorists have strongly emphasized their connection (e.g., [22], [109]), during the 20th century the fields of motivation and emotion psychology have to a large degree developed independently. As a consequence, some of the best-known psychological theories of motivation and decision making, such as the theory of planned behavior [2] or prospect theory [84], do not consider emotions, at least not explicitly. Although it can be argued that even in these formally affect-free theories, emotions are implicitly considered as a possible source of utility (i.e., as hedonistic desires, as explained in Section 2.2.4; see [10], [42]), the explicit recognition of emotions as separate variables in modern motivation and decision theories is of relatively recent origin (e.g., [113], [180], [203]). Likewise, 20th century emotion theorists have typically paid little attention to theories of motivation and decision making. That is, in developing their theories of the generation and nature, and even the effects of emotions, these emotion theorists have typically not tried to explain (at least not in detail) how emotions fit into the mechanisms of goal and action selection described by existing theories of motivation [145].

To further clarify the relation between emotion and motivation, we briefly consider the question from the perspective of motivation theory. It can be argued that nearly all current theories of motivation and action generation are variants of a singly basic theory, the belief-desire theory of action (for a selective sample of the relevant literature, see [29], [41], [59], [60], [66], [112], [172], [183], [202], [206]). The basic assumption of the belief-desire theory of action is that actions—or more precisely, intentions to act, which then cause actions under favorable circumstances—are products of beliefs about one's current state and means-ends relations, and desires for the ends.[12] Philosophical proponents of the theory have argued that to produce adaptive actions, exactly these two kinds of representational states, the one informational and the other motivational, are needed (e.g., [183]): Without beliefs (about the current state of the world and about ways to achieve goals), desires would be blind; without desires, beliefs would be inert. These considerations suggest that—contrary to the claims of some emotion theorists (e.g., [22] and possibly [43])—emotions are *not indispensable* for the generation of adaptive actions; even though affect-free actions may well be overall less adaptive than actions that are also informed by emotions (see Section 2.2.4). Support for this conclusion might be seen in the fact that it is possible to develop workable theories of motivation and decision making without explicitly considering emotions (e.g., [2],

10. It has also been proposed that the main evolutionary function of emotions is to coordinate the subsystems of the organism into an adaptive whole-organism response to evolutionarily significant challenges and opportunities (e.g., [110], [127], [166], [196]). However, the available empirical evidence suggests that the components of emotional syndromes are in fact only loosely associated (e.g., [17]) and the observed clustering can probably be explained as the result of the three described functional effects.

Other documented effects of (some) emotions comprise the priming of emotion-congruent material in memory and the biasing of information processing in emotion-congruent ways (e.g., [7], [23], [170]); as well as physiological activation. Whereas physiological activation can be given a reasonably plausible functional explanation (it served to prepare fight/flight reactions in evolutionary predicaments; [35], [166]), the functionality of the biasing effects of emotion on memory and information processing is debated (for one possible functional explanation, see [158]).

11. According to Weiner [202], motivation psychology is concerned with explaining why humans behave as they do. If one accepts that human behaviors are for the most part goal-directed, this means that motivation psychology is centrally concerned with explicating the processes that generate goal-directed behavior [42]. Thus, a theory of motivation is essentially a theory of the mechanisms of action and goal generation, including a theory of basic goals or fundamental desires.

12. Decision theories (including subjective expected utility theory, prospect theory, and diverse psychological expectancy-value theories such as the theory of planned behavior) can be regarded as different attempts to quantify a generalized version of the belief-desire theory of action [145]. However, the relation between *specific versions* of belief-desire theory, such as BDI theory, and decision theory is not straightforward (e.g., [47]).

[84]). Support might also be seen in the fact that it has been possible to create artificial agents that show reasonably adaptive goal-directed behavior even though their actions are not explicitly influenced by internal states corresponding to emotions (e.g., BDI agents; [142]; see Section 4.2).[13] By contrast, the generation of intentional actions is simply unthinkable without motivational mechanisms—mechanisms of goal selection and action generation that invoke desires and beliefs (see also, [199]). In line with this, it appears that to the degree that the effects of emotions on intentional action have been discussed in some detail by emotion theorists, they presuppose for the most part, at least implicitly, a traditional motivation or decision theory (e.g., [9], [43], [62], [92]).

Perhaps these considerations may be taken to suggest the following: First, the central evolutionary function of emotions is to improve the generation of adaptive action beyond what is possible with beliefs and desires alone. Second, emotions achieve this by influencing, in one way or other, the motivational (belief-desire) machinery that proximately controls actions. This may be what unites the above-described informational, attentional and motivational effects of emotions.

2.2.6 Formal Tools for Representing Emotion Theories

Although informal analyses of emotion theories can go a long way toward clarifying and systematizing them, even the reconstructed theories will usually not be sufficiently precise and complete to be directly implemented as computational models. Of course, increasing the precision and completeness of a theory is precisely one goal of creating a computational model of that theory (e.g., [61], [192]). However, at this point in the history of computational modeling, the *direct* implementation of an emotion theory as a computer program is usually not the best way to go—at least if this means to program everything from scratch. For one reason, this is a cumbersome way to proceed, as it requires to invent or at least program much anew; for another reason, too many potentially irrelevant implementation choices have to be made along the way. In this and the following section, we consider two alternatives: 1) Formalizing emotion theories in a precise but implementation-independent formal language; and 2) implementing emotion theories in a (comparatively) general-purpose cognitive or agent architecture.

3 Theory Formalizations: Set Theory and Agent Logics

The first proposal consists of translating a psychological emotion theory (or selected assumptions of one or more emotion theories, as discussed in Section 2.2), as an intermediate step, into a precise, but implementation-independent language. Such a formalization has two main benefits (e.g., [74]). The first coincides with the general benefit of theory formalization. This benefit derives to a large extent from the fact that the formalization of a theory is usually much more than a simple *translation* of the theory into a formal language; it is simultaneously a process of *explicating* the theory: resolving ambiguities and possible inconsistencies, uncovering hidden assumptions, adding missing assumptions, and the like (e.g., [205]). The result, ideally, is an improved version of the original theory: a version that is precise, consistent, complete, and concise. In addition, formalizations provide improved opportunities for comparing and evaluating the assumptions of different theories described in the same format, they facilitate the derivation of predictions, and they can raise new questions that may even suggest additional empirical research.

However, it can be argued that the general benefits of formalization are shared by computational models (which can be regarded as formalizations of substantive theories in a programming language; e.g., [108]).[14] What are the special benefits of first creating a formal specification of a psychological emotion theory, that is then implemented as a computational model? Perhaps the main benefit is that such a formalization yields a precise model of the psychological theory but avoids a commitment to potentially irrelevant implementation details (that are a necessary part of a computational model). That is, only assumptions regarded as essential for the emotion theory in question need to be considered in the formalization, whereas computational details—even important ones such as the structure of the agent architecture—can be ignored if they are considered irrelevant [74]. As a consequence, less effort is needed to verify that a formal model is a correct representation of an emotion theory, than would be required to verify a computational model. In addition, a precise and unambiguous formal version of an emotion theory facilitates its subsequent implementation as a computational model (see also, [45]). Finally, like the source theories on which they are based, formalized theories still allow to deduce implications (make predictions) *directly*, whereas computational models need to be *run* to generate predictions. Because of their special virtues, theory formalizations in implementation-independent languages can also be useful for comparing existing computational models of emotions.

3.1 Set-Theoretical Formalization of Emotion Theories

Broekens et al. [32] propose to use *set theory* as a language for formalizing psychological emotion theories (see also, [147]). Set theory recommends itself as a formalism for the precise reconstruction of theories because of its expressiveness; in particular, (nearly) all mathematical concepts can be defined in terms of set-theoretical concepts [Bourbaki, 28].[15] Choosing set theory as the formalization language therefore makes all of classical mathematics available for theory reconstruction and essentially amounts to the mathematization of a

13. As mentioned, the belief-desire theory of action and its quantitative version, decision theory, do not exclude an influence of emotions (in the form of hedonistic desires). Our argument here is that these theories do not *require* emotions; in particular, nothing in these theories enforces a hedonistic interpretation of desires. Actions require desires for goals, but as already stressed by McDougall [109], the goals need not be hedonistic.

14. More precisely, one should say that *certain parts* of computational models (those corresponding to the theory) algorithmically *implement*, or *realize*, substantive theories (e.g., [102]).

15. Note that the Bourbaki approach uses informal rather than formal (in the sense of axiomatic) set theory; hence, this approach can be described as "informal formalization" [184].

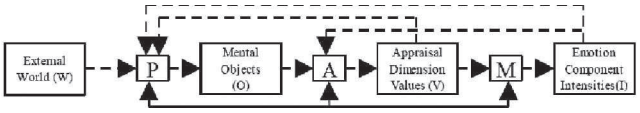

Fig. 1. Basic framework for a set-theoretical formalization of appraisal theories [32].

theory (by which we understand the construction of a mathematical, although not necessarily a quantitative model). A set-theoretical formalism is also well suited to represent computational theories (as well as psychological theories interpreted as information-processing theories), because a common way of describing computational processes is as mathematical functions that map input to output information. This formal representation translates fairly directly into familiar graphical descriptions of systems, such as the block diagrams used in cybernetics and computer science and associated methods such as functional decomposition, and is also closely related to causal modeling formalisms used in psychology (e.g., [26]). These parallels can be exploited to depict computational functions in a more accessible graphical format (e.g., [32]) and conversely, to translate graphical theory proposals into set-theoretical language.

For example, Broekens et al. [32] break up the emotion generation process assumed by cognitive appraisal theories into three linked functions (see Fig. 1): *Perceptual processes* map external objects into percepts (or mental objects); *appraisal processes* map mental objects into different appraisal dimension values; and *mediating processes* map appraisals into components of emotional reactions (e.g., feelings, expressions, actions). Using this basic functional (or process) decomposition, Broekens et al. then propose set-theoretical formalizations of two specific appraisal theories, Scherer's [166] stimulus evaluation check theory, and Smith and Kirby's [181] appraisal detector theory.

The set-theoretical formalization of emotion theories used by Broekens et al. [32] ties in naturally with a still more rigorous set-theoretical formalism for theory reconstruction, developed as part of the so-called *structuralist program* in the philosophy of science (e.g., [14], [15], [204]). Essentially, this formalism consists of set theory (including whatever parts of mathematics are needed) enriched with a set of precisely defined special concepts that describe common, recurring components of empirical theories. According to structuralism, empirical theories are best viewed, from the systematic perspective, as set-theoretic structures composed of so-called *theory-elements*. At minimum, a structuralist reconstruction of a theory consists of a single theory element, but frequently it consist of several hierarchically connected theory-elements, called *theory-nets*. In the typical case, theory-nets comprise a basic element that represents the fundamental assumptions of a theory, plus a number of specialized elements that result from the basic element through the more precise specification of existing or the addition of further assumptions. In this way, theory nets allow to represent both refinements and different elaborations of a common core idea. In addition, theory elements belonging to different theory nets can be connected through *intertheoretical links*. Links are used to formally express the fact that empirical theories typically presuppose other theories (e.g., theories of measurement of their basic quantities). The basic building blocks of theories, the theory elements, have a precisely defined internal structure: They are defined as tuples $<\mathbf{K}(\mathbf{T}), \mathbf{I}(\mathbf{T})>$, where $\mathbf{K}(\mathbf{T})$ is the mathematical core of the theory, and $\mathbf{I}(\mathbf{T})$ is a set of intended applications of the theory core. $\mathbf{K}(\mathbf{T})$ consists in particular (but not exclusively) of the classes of potential models ($\mathbf{M}_\mathbf{p}$) and the class of models ($\mathbf{M}$) of the theory. Potential models are all entities (set-theoretical structures, or systems) that can be described in terms of the theory's basic concepts, and for which it is therefore meaningful to ask whether they fulfill the substantive laws of the theory; whereas models are potential models that also fulfill these laws.

To use an elementary example, consider a structuralist reconstruction of a miniature quantitative belief-desire theory of hope and fear (e.g., [151]). The class of potential models of the basic element of this theory could be defined as the set of systems $\mathbf{x} = <PROP, TIMES, b, d, hope, fear>$ consisting of a set of propositions $PROP = \{p_1, p_2, \ldots, p_n\}$, a set of time points $TIMES = \{t_1, t_2, \ldots, t_m\}$, a belief function b: $PROP \times TIMES \rightarrow$ [0, 1], a desire function d: $PROP \times TIMES \rightarrow \mathbb{R}$ (the set of real numbers), and functions for *hope* and *fear*, both of which are also defined on $PROP \times TIMES \rightarrow \mathbb{R}$. A system $\mathbf{x}$ belongs to the class of models of this mini-theory of hope and fear if $\mathbf{x}$ is a potential model of the theory and in addition fulfils the following two laws (which are possible specializations of the more general functions proposed in [151]; see also, [95]):

$$\begin{aligned} hope(p,t) &= b(p,t) \times d(p,t) \text{ if } 0 < b(p,t) \\ &< 1 \text{ and } d(p,t) > 0; \text{else } 0, \\ fear(p,t) &= |b(p,t) \times d(p,t)| \text{ if } 0 < b(p,t) \\ &< 1 \text{ and } d(p,t) < 0; \text{else } 0. \end{aligned}$$

Because the definitions of set-theoretic predicates never yield a sufficient characterization of an empirical theory, the theory's range of application is determined by describing a set of intended applications $\mathbf{I}(\mathbf{T})$. This is done by 1) explicitly listing a few real systems, called *paradigms*, and declaring them to be intended systems, or applications of the theory core; and by 2) declaring all other systems to be intended applications if they are sufficiently similar to the paradigmatic cases. For example, one paradigmatic application of hope-fear theory would be the experiences of hope and fear of a person awaiting a possible gain (hope) or loss (fear) in a two-outcome lottery where the other outcome is zero [156]; whereas another intended application might be the hope and fear experiences of gamers at certain stages in a computer game [69].

The viability of the structuralist approach to theory reconstruction is documented by the fact that at least 50 reconstructions of different theories from diverse disciplines have been published (see, e.g., [15], [204]), including several theories from psychology. Westmeyer [205], therefore, recommended the structuralist approach as a general methodology for theoretical psychology; while Balzer and Moulines [13] recommended the structuralist formalism of theory representation (specifically to computer scientists) as

a good tool for representing scientific knowledge in a general context.[16] Furthermore, the structuralist view of theories has inspired a simulation language for social science applications (MIMOSE; now obsolete) that facilitates the transformation of a structuralist theory reconstruction into an executable program ([197]; see [198] for other simulation languages that show a similar correspondence to set-theoretic formalizations[17]). Although so far only one serious structuralist reconstruction of an emotion theory (the 3D theory of emotion proposed by Wundt [207]) seems to have been attempted [146], the sketch of the belief-desire model of hope and fear given above and the set-theoretic formalizations of the two emotion theories proposed by Broekens et al. [32] suggest that rigorous structuralist reconstructions of other emotion theories are certainly feasible. Importantly, as mentioned above, structuralist reconstructions of psychological theories are not restricted to intentional-level theories [51], but extend to design-level (computational) theories. Hence, the structuralist idiom could also be used for an abstract description of computational emotion models realized in cognitive and agent architectures such as Soar (e.g., [68], [104]; Section 4.1), the belief-desire-intention (BDI) architecture (e.g., [20]; Section 4.2), or affective agent architectures (e.g., [77]; Section 4.3). In fact, a structuralist reconstruction of a precursor of ACT-R (ACT*, [4]) has been presented by Heise [72].

3.2 Formalizing Emotion Theories in Agent Logics

Even though Balzer and Moulines [13] recommended the set-theoretic formalization of theories for representing scientific knowledge in a general context, they acknowledged that more restrictive formalisms may be more useful for specific applications. In particular, applied computer scientists who seek to develop a computational model of an emotion theory may prefer a formal specification language that, while less expressive than set theory, provides them with more structure, to facilitate the transformation from specification to implementation. One particularly interesting alternative possibility of formalizing emotion theories—at least intentional-level, cognitive emotion theories (see Section 2.2.3)—is to represent them in a *logic language*. During the last 20 years, several logic languages have been developed for modeling cognitive autonomous agents that are suited for this purpose. Most of these so-called *agent logics* belong to the class of (broadly conceived) *belief-desire-intention logics*, that describe autonomous agents on the intentional level [51] in terms of beliefs, desires (goals), intentions (roughly, the goals that the agent has committed to) and possibly other related attitudes, such as capabilities (see, e.g., [54], [199], [206]).

More precisely, BDI logics are formal logic languages that arise from the combination of several modal logics: a temporal or a dynamic logic used to capture the dynamic nature of agents, and logics for the mental states of belief, desire and intention. Each of the modal operators is given a precise syntactical definition in terms of a set of axioms, and a precise semantics in terms of possible worlds models (Kripke structures; see, for example, [81]). The axioms for belief, desire and intention and their interaction encapsulate principles of thinking (belief-formation) and deciding for an idealized rational agent that reasons qualitatively [206]. Put differently, these axioms encapsulate a qualitative version of rational decision theory [47], whose basic decision principle is simply that rational agents intend to do what they believe will lead to what they desire [21], [138]. To this, BDI theory adds the important idea that rational decision makers need to show some commitment to their intentions [29].

As mentioned in Section 2.2.3, most (at least most "higher") emotions are thought to depend on beliefs and—directly or indirectly—on desires. Therefore, BDI logics are natural candidates for formalizing theories of (these) emotions. A second reason for formalizing emotion theories in BDI logics is that these logics already contain axioms for decision making and hence describe a workable, autonomous goal-directed agent. Formalizing an emotion theory in a BDI logic thus amounts to introducing emotion concepts into a formal theory of rational agency, or a formal qualitative decision theory. This approach facilitates the modeling of the effects of emotion on action (e.g., [46]) and the systematic investigation of the benefits of emotion mechanisms for goal-directed rational agents (e.g., [45], [97], [99]).

Although semiformal belief-desire analyses of some emotions have been proposed by philosophers at least since the 1970s (e.g., [48], [50], [70], [171]), rigorous formalizations of belief-desire analyses of emotions in agent logics are of relatively recent origin. Most of these formalizations focus on the cognitive and motivational preconditions (or, depending on viewpoint, the components; see [70]) of emotions (e.g., [1], [36], [114], [133], [186], [187]). For example, Castelfranchi and Lorini ([36]; see [96] for an update) formalized the belief-desire preconditions of a set of emotions related to expectations (hope, fear, disappointment, and relief) using one of the first BDI logics, proposed by Cohen and Levesque [40]. Meyer [114] used KARO, an agent logic based on a dynamic logic augmented with BDI-type modal operators [200] to formalize the belief-desire preconditions of four basic emotions (happiness, sadness, anger, and fear) proposed in Oatley and Johnson-Laird's [127] theory of emotion, another belief-desire theory (see also [126]). More recently, Steunebrink et al. [187] used KARO to formalize the cognitive-motivational preconditions of the 22 emotions considered in the OCC theory [129]. Another formalization of the OCC theory, using an extended version of the Cohen-Levesque logic, was proposed by Adam et al. [1]. As mentioned in Section 2.2.3, these formalizations involve a reinterpretation of the OCC theory as a belief-desire theory of emotion.

In addition to formalizing hypotheses about emotion generation, recent work has formalized some of the proposed effects of emotion (see Section 2.2.4), specifically their effects of actions [187], [46], including expressive speech acts [71].

The main motivation behind these logic formalizations of emotion theories has been to provide rigorous formal

16. According to the authors, "there seems to be a fact of the matter that requires a set-theoretic representation of scientific knowledge, all weaker (and logically nicer) formal systems [being] too poor in expressive power or otherwise too restrictive for the working scientist" ([13, p. 9]).

17. Thanks are due to Klaus Troitzsch for providing this information.

specifications for the implementation of emotions in autonomous agents. However, as mentioned, these formalizations can also be useful for psychologists and philosophers of emotion: Apart from providing a concise summary of a psychological or philosophical source theory, they can help to clarify and disambiguate assumptions of the source theory, reveal problematic implications that may otherwise remain hidden or go unnoticed, facilitate the derivation of consequences, and raise new questions [1].[18] Several examples can be cited to support these claims:

1. Steunebrink et al. [186] as well as Adam et al. [1] found that their logical reconstructions of the OCC theory revealed a number of ambiguities in that theory, such as whether event-based emotions are caused by perceived events or the believed consequences of events.
2. Adam et al.'s [1] reconstruction of the OCC theory led them to propose, for example, that an agent is disappointed about a state of affairs not-p if it believes not-p to be the case, desired p to occur, and believes that he previously expected p. Hence, agents need to form a metacognition (a belief about an own belief) to experience disappointment. Although this requirement may be unproblematic for artificial agents, at least in principle (though not for some concrete realizations of BDI agents; for example, [73]), it could be problematic for humans [151]. Adam et al.'s formalization makes this assumption transparent and thereby presents it for discussion. In fact, Castelfranchi and Lorini [36] and Steunebrink et al. [187] have proposed alternative analyses of disappointment that avoid postulating metacognitions. For example, Castelfranchi and Lorini assume that an agent is disappointed about not-p if it previously hoped that p will occur and then comes to believe not-p (see also [151]).
3. The two formalizations of the OCC theory by Adam et al. [1] and Steunebrink et al. [187] also differ in several other respects (see [187] for a discussion) that could stimulate discussion and possibly even empirical research.
4. The logical formalizations bring into sharp focus the question of the nature of emotions in belief-desire theory. In the majority of the existing formalizations (the exception is [187]), emotions are *defined* in terms of beliefs and desires—for example, $joy(p) =^{\mathrm{df}} bel(p)$ & $des(p)$—which might suggest that emotions are to be *equated* with particular belief-desire structures. This has indeed been proposed by some belief-desire theorists of emotion (e.g., [105]; see also [70]). However, this is neither the only nor necessarily the most plausible version of the belief-desire theory of emotion (see [153] for a detailed discussion). In fact, most authors make clear that they do not intend their formalizations as implying a reduction of emotions to beliefs and desires, or at least do not wish to commit to this theory of the nature of emotion. Rather, the formalization is meant to express, or at least to be also compatible with, the idea that beliefs and desires are necessary, and under normal circumstances also sufficient, *causes* of emotions.

Let us now consider two possible, general objections to formalizations of emotion theories in BDI logics. The first is that BDI logics *are not expressive enough* to accommodate all assumptions of psychological emotion theories. Three concrete forms of this objection are the following: 1) The above-mentioned formalizations are purely qualitative; however, emotions differ not only in quality but also in intensity, and intensity is important for explaining certain effects of emotions [64]. To account for emotion intensity, quantitative concepts of belief and desire are needed. 2) Even on the qualitative level, BDI logics may not be rich enough to formalize the cognitive-motivational preconditions of all emotions. For example, it has been argued that feeling guilty about p presupposes not only the belief that p is the case and the desire that p should not have occurred, but also the belief that one is responsible for p (e.g., [129], [203]). 3) Although BDI logics go beyond traditional "static" decision and action theories in considering the dynamics of mental states and actions, they are still restricted to the intentional level of system analysis [51]; they do not address the underlying representations and computational processes.

These objections can be answered as follows: As to the first objection (BDI emotion models are qualitative), it can be argued that for one thing, precise qualitative formulations are no less important than quantitative ones (e.g., [129]). For another thing, it is possible to extend BDI logics to account for emotion intensity. For example, Lorini [96], [46] proposed to combine a dynamic logic of knowledge with a logic for graded beliefs and desires for this purpose. Similarly, Steunebrink et al. [185] have incorporated quantitative aspects of emotion into their KARO model.

The second objection can be answered similarly. For example, to explicate the concept of responsibility and related concepts such as preventing and causing, Lorini and Schwarzentruber [98] have proposed to combine BDI logic with STIT logic (a logic of "seeing to it that").

As to the third objection (logical emotion models are restricted to the intentional level of system analysis), this restriction is deliberate and does not preclude combining the logical models with design-level assumptions, as are in fact provided by BDI architectures (see Section 4.2). On the contrary, as mentioned, because of their greater precision, logical BDI models can *assist* the creation of computational models. Nevertheless, at this juncture, the interests of psychologists and computer scientists may partly diverge. From the perspective of computer scientists, at least those pursuing applied goals, a specification of an emotion theory that abstracts completely from questions concerning representational formats and computational mechanisms may be regarded as a benefit. In contrast, as mentioned in Section 2, psychologists adhering to the computational paradigm are

18. The fact that a variety of different BDI-type logics are available to choose from should probably be regarded as an asset rather than as a disadvantage. First, it means that one need not commit to a particular logic; different routes can be tried and compared. Second, the existence of different logics suggests that basic conceptual issues need to be, and can be, explored while formalizing emotion theories.

equally interested in certain design-level questions, although not in all (e.g., [68]; see also [151], [178]).

A second general objection that could be raised against formalizing emotion theories in BDI logics is that the rationality assumptions incorporated in the usual axioms for belief, desire and intention imply that the logical BDI models can only be regarded as strong idealizations of human belief-formation and decision making. For example, standard BDI logics assume that agents believe and desire all logical implications of their beliefs and desires, respectively, and are also perfectly aware of the beliefs and desires that they have (e.g., [58]). These are not plausible assumptions to make for real agents with limited memory, limited processing time and (most likely) limited metacognitive abilities. Because these rationality assumptions carry over to beliefs and desires as presuppositions of emotion, the logical models of "emotion-enhanced" BDI agents can likewise be regarded only as strong idealizations of human emotion mechanisms.

To some degree, this criticism can be met by formulating at least the emotion axioms in ways that do not demand more from agents than is necessary and psychologically plausible (see [36], [187]). Beyond that, attempts have been made to develop logics that describe "less than ideal" agents (e.g., resource-bounded epistemic logics, [115]). However, this project has its own problems; for example, Duc [56] argues that by axiomatically restricting the rationality of agents, nothing is left that can account for their rationality any more. From the perspective of psychology, two additional considerations may be more important: First, it should be remembered that all theories are idealizations and that even strong idealizations can be quite useful (consider, for example, the case of game theory; e.g., [34]). In particular, the strong rationality assumptions of BDI logic can be mentally set aside when the logic is used as a tool to clarify hypotheses (see [41]); for this purpose, what mainly counts is the increased precision provided by the formal language. Second, *computer implementations* of BDI agents (as opposed to their logical descriptions) have to operate under time and memory constraints like humans. As a consequence, they become *boundedly rational agents*, rational agents that behave as well as possible given their limited resources and their constraints (e.g., [169]). To obtain more realistic models of motivational and emotional processes, psychologists can, therefore, focus on the BDI architectures (see Section 4.2).

4 Modeling Emotions Using Cognitive and Agent Architectures

Another proposal for bridging the gap between psychological emotion theories and computational emotion modeling, as well as for increasing the comparability of different computational emotion models, is to encourage the use of (suitably extended) *general-purpose cognitive and agent architectures* to model emotion theories. Cognitive architectures and agent architectures are theories of the *basic design-level structure* [51] of cognitive systems or agents, respectively: theories that describe their domain-independent, basic computational structures (e.g., memories, representational systems) and processes (e.g., a repeated sense-think/decide-act cycle) (e.g., [6], [88], [164], [192]). From the perspective of the user (modeler), an architecture appears as the "hardwired" part of an agent that cannot be changed (at least not easily).

Architectures are typically used to facilitate the creation of computational models of specific psychological phenomena of interest, such as the performance on a particular task (e.g., in the case of emotion research, the performance on the Iowa gambling task; [188]). The main advantage of using an architecture for this purpose, compared to creating a computational model from scratch, is that the architecture takes care of numerous implementation choices that otherwise would have to be made. Thereby, it relieves the modeler of the need to invent everything anew and helps—similar to the formal languages discussed in Section 3, but in a different way—to solve the problem of irrelevant design specifications: Whereas in the formal models, design assumptions considered irrelevant are *omitted*, in an architecture, they are *assigned plausible default values*. By letting the architecture take care of design decisions, the modeler implicitly accepts that the respective assumptions of the architecture are correct, or at least not too far off the mark in ways that matter to the phenomenon of interest. If one is willing to make this assumption, the use of an architecture can greatly simplify the creation of a computational emotion model [108]. Furthermore, to the degree that architectures are becoming reasonably easy to use even for nonprofessional programmers, they could significantly lower the threshold for more psychologists becoming interested in, and even engaging themselves in the computational modeling of emotions. In addition, the research efforts of affective computing scientists could be focused, and competing proposals more easily compared, by modeling emotions in otherwise well-understood architectures.

However, the role of architectures in computational modeling extends beyond these practical benefits: Architectures can also play important *theoretical roles*. In particular, to the degree that an architecture is well-anchored in theory and research and hence contains many at least approximately correct psychological assumptions, the computational distinctions incorporated in the architecture (e.g., between deliberative and automatic processing) can be exploited to clarify similar distinctions proposed in a psychological process theory of emotion (see, e.g., [68], and Section 4.1). Sloman [178] even argues that an understanding of emotions can *only* be attained by considering the role of emotions in cognitive architectures. As Sloman puts it: "A proper analysis of the concept of an "affective" state or process must be based on a more general theory of the coarse-grained architecture of mind" ([178, p. 233]). This said, it should be acknowledged that the fixed computational structure provided by a given architecture can also be a disadvantage: namely, if it constrains what one wants to model in crucial respects. This consideration is particularly important in the computational modeling of emotion, because emotions may be implicated in basic psychological processes at the architectural level (see Section 2.2.4; see also [151]).

General overviews of cognitive and agent architectures are provided by Chong et al. [38], Langley et al. [90], Vernon et al. [201], and Zacharias et al. [208]; for overviews of emotional agent architectures, see Hudlicka [78] and Marsella et al. [108]. For the purpose of our review, we

have grouped the (comparatively) general-purpose cognitive and agent architectures that can be and in fact have been used for emotion modeling into 1) by now "classical" cognitive architectures such as Soar or ACT-R (see Section 4.1); 2) autonomous agent architectures developed since the 1990s in artificial intelligence, the most widely used probably being the BDI architecture (see Section 4.2); and 3) (comparatively) general-purpose architectures specifically designed for the creation of affective agents, such as FAtiMA Modular and MAMID (see Section 4.3). Despite important differences between these architectures, they share some basic similarities. In particular, all of them can be taken to describe goal-based (or goal-directed) agents [57] rather than (merely) reactive agents (although reactive behaviors can also be modeled); and all use symbolic internal representations to represent the contents of the agent's mental states. Furthermore, none of these architectures does anything on its own; it is only when they are supplied with data (knowledge/beliefs, goals) and placed in a suitable task environment that they "come to life" and generate cognitive processes and behavior.

The main difference between the affective agent architectures and the more traditional cognitive and agent architectures is that the latter do not, in their original form, consider emotions. Nevertheless, these architectures have been used to model both processes of emotion elicitation ([68]; using Soar) and the effects of emotion on action and task performance (see again [68] for Soar, and, e.g., [160] for ACT-R). This has been possible because both Soar and ACT-R contain, at least implicitly and in an approximate fashion, central cognitive and motivational constructs needed to account for emotions (see Section 2)—constructs relating to the beliefs (knowledge) and the motivation (desires, goals) of agents (for more detailed discussions, see [57], [65], [107]).

4.1 Modeling Emotions in Cognitive Architectures

Classical cognitive architectures such as Soar and ACT-R are "echos of the original goal [of AI] of creating an intelligent machine faithful to human intelligence" (Taatgen and Anderson [193, p. 170]). At the same time, they are attempts at theoretical unification in the field of cognitive psychology (again [193, p. 170]).

4.1.1 Modeling Emotions in Soar

One popular cognitive architecture, that has also been used to concretize emotion theories, is Soar [88], [123]. Soar is a general cognitive architecture for human cognition realized as a production system, the basic operation of which consists of matching the contents of working memory to the precondition of rules (see [164]). Following the classic work of Newell and Simon [124] on the *General Problem Solver*, the basic assumption of Soar is that intelligent behavior is a form of problem solving. In Soar, tasks are represented as problem spaces and problem solving consist of a search (involving both automatic and deliberative processes) for sequences of operators that will reach a goal state in problem space. At each step in this search, Soar rapidly and in parallel recruits knowledge from long-term memory that is relevant to the current situation. After this knowledge recruitment processes reaches quiescence, the decision procedure executes a single most-preferred operator, records its consequences to working memory or the outside environment, and then recruits knowledge relevant to this new state. If a problem arises due to the fact that certain knowledge is lacking, resolving this problem becomes a subgoal of the original goal. Soar is also capable of learning about how subgoals were reached, and what the consequences of operators are. It may be noted that considered from the intentional stance [51], the operation of Soar is essentially a form of practical reasoning ([29]; see also [65]): Goals are desired states to be reached by actions; operators correspond to possible actions, and knowledge about the effects of operators correspond to means-ends beliefs.

Soar was not designed to model emotions; nevertheless, two different computational emotion models have been built upon Soar [68], [104]. These two models represent the two principal alternative paths available to model emotions in Soar and similar cognitive architectures, and they also illustrate how the theoretical assumptions embedded in an architecture can influence modeling choices.

EMA, proposed by Gratch and Marsella ([68]; see also [107]), implements the cognitive process theory of emotion proposed by Smith and Lazarus [182]. In EMA, the basic architecture of Soar is left largely untouched, by implementing appraisal processes as production rules encoded in Soar's long-term memory. In doing this, EMA exploits Soar's theoretical distinction between deliberative and automatic processes to concretize a similar distinction proposed by Smith and Lazarus between deliberative (conscious) and automatic appraisal processes: Appraisals are realized in Soar as "elaboration" productions that fire rapidly and in parallel, and that bias sequential operator decisions. Thus, *by following Soar's own theoretical assumptions*, EMA achieves a tight and cyclical coupling between automatic and deliberative appraisal processes in a way that can be regarded as a computational specification of Smith and Lazarus's process theory of emotion.

In contrast to EMA, Marinier et al.'s [104] PEACTIDM model introduces emotions into Soar by changing Soar's core architecture: Inspired by Scherer's [166] theory of sequential appraisal checks, PEACTIDM attempts to capture the sequential nature of appraisals posited by in that theory by adding sub-stages to Soar's decision procedure. Specific appraisals are encoded into the architecture and are associated with these processing stages. As a result, PEACTIDM, like EMA, computationally implements the assumption of a tight coupling between automatic and deliberative appraisal processes; however, it takes a very different path to realize this idea.

4.1.2 Modeling Emotions in ACT-R

ACT-R [5], [6] is another influential cognitive architecture that shares many similarities with Soar; in particular, like Soar, ACT-R is a production system with goals that operates on symbolic representations. Also like Soar, the basic operation of ACT-R is to achieve goals by applying suitable procedures (actions). The differences between Soar and ACT-R concern mainly a variety of design features that together are intended to make ACT-R more psychologically plausible than Soar in its original form ([6]; for a discussion of some differences, see also [83]). These design features include an explicit distinction between a procedural and a declarative memory and multiple modality-specific memory

buffers.[19] A particular strength of ACT-R is its well-developed model of long-term memory, in which information is represented as networks and is activated by a subsymbolic process of spreading activation (in fact, ACT-R developed out of a theory of human associative memory). In this sense, ACT-R can be said to represent an integration of symbolic and subsymbolic mechanisms [160].

Although ACT-R has so far only been used to a limited extent to model emotions, it deserves particular attention because it is probably the best known cognitive architecture in psychology and has been used to create precise computational models of many psychological theories, including several models of decision-making (e.g., [103]). Ritter et al. [160] used ACT-R to demonstrate two general ways of how some of the effects of emotions can be modeled in cognitive architectures, namely 1) by varying the content of the (procedural or declarative) knowledge of the agent; and 2) by varying certain architectural parameters (as proposed by Hudlicka [76]; see also [77], [79]). The first approach was used to model the effects of worry on performance, by including a task-irrelevant production in ACT-R that could fire at any time, simulating the effects of distracting thoughts on working memory. The second approach was used to model the effects of threat versus challenge appraisals of a stressful task on performance, by varying the level of randomness in the conflict resolution process (which determines which production rule will fire when more than one rule matches the goals of the system). Other noteworthy applications of ACT-R to emotion modeling are Cochran et al.'s [39] model of the impact of arousal on memory, and Stocco et al.'s [188] model of the Iowa gambling task, the main experimental paradigm used to test Damasio's [43] somatic marker hypothesis of the effects of emotion on decision making. Although *emotion elicitation processes* do not seem to have been modeled in ACT-R, the similarities of this architecture to Soar suggest that these processes could be considered in one of two ways in ACT-R: 1) by modeling appraisals as production rules (analogous to the approach taken by Gratch and Marsella [68] in Soar); or by extending the basic ACT-R architecture (analogous to [104]).

4.2 Modeling Emotions in BDI Agent Architectures

In Section 3.2, we discussed logical models of belief-desire-intention agents and noted that, among other uses, such models can aid the creation of corresponding computational models. Here, we discuss the associated computational BDI architecture (e.g., [30], [142], [143]). To recall, logical BDI models are formalized theories of qualitative rational decision making [29]. However, although these models are precise renderings of informal intentional-level [51] theories, they remain silent about the design level: They make no assumptions about how beliefs, desires and intentions, and the mental operations performed with them, are computationally implemented. BDI architectures, in contrast, (also) address these questions: They are pieces of software that implement the abstract principles of BDI agents.

The BDI architecture, which exists in several variants, has become one of the most frequently used software architectures for autonomous intelligent agents. The central assumption incorporated in BDI agents is that actions are generated by a process of practical reasoning [29] that comprises two computational steps. In the first step, a set of desires (goals) is selected; in the second step, the agent determines how these goals can be achieved by means of the available actions or plans [206]. Characteristic for BDI agents, furthermore, is the separation of the process of *deliberation* from the process of *execution* of currently active plans. This allows BDI agents to balance the time spent on goal and plan selection (choosing what to do) and executing plans (acting), which seems to be essential for adaptive action in resource-bounded agents [29].

Arguments in favor of BDI architectures as blueprints for autonomous agents are their foundation in philosophical (and psychological) theory; the existence of logical models that allow to define, and reason about, BDI agents in a precise manner (see Section 3.2); and the existence of several software systems for programming BDI agents that are close to these logical specifications (e.g., JAM [75]; JADEX [136]; 2APL [44]; GOAL [73]). Compared to cognitive architectures such as Soar and ACT-R, BDI architectures are more abstract; for example, they do not require a commitment to specific memory types. Depending on one's modeling goals, this can be an advantage or a disadvantage (e.g., it is a disadvantage if one's interest is to model the effects of emotion on memory). Most BDI systems execute a fixed perception-deliberation-action (sense-think-act) execution cycle comprising the following steps:

1. perception of events;
2. updating of the belief and desire (goal) base (adding new beliefs and desires, dropping false beliefs and satisfied or impossible desires);
3. generating options (plans) (deliberating);
4. choosing the most suitable plan (generating an intention); and
5. executing intentions (acting).

The basic BDI architecture has been extended in several ways, either by modifying the agent's internal processes or by extending the BDI model to the social environment (communication, multi-agent BDI systems). "Internal" extensions of the BDI architecture comprise the inclusion of learning mechanisms, the consideration of personality differences (e.g., [125]), and the inclusion of emotions.

BDI architectures are in principle well-suited to create computational models of emotions given that most (at least most "higher") emotions are thought to depend on beliefs and—directly or indirectly—on desires (see Section 2.2.3; see also [74]). Furthermore, the formalization of emotion theories in BDI logics discussed in Section 3.2 should facilitate their implementation in BDI software systems. Indeed, several attempts have already been made to incorporate (aspects of) psychological emotion theories into BDI systems (e.g., [20], [140], [141]). For example, Rank [140], [141], has extended JAM agents [75] by including an appraisal mechanism that compares perceived facts concerning own actions and those of other agents (success, failure, attempt) with the agent's goals and computes positive or negative "relevance values" on the basis of the utility of the affected top-level goal and the match or mismatch of the action-relevant facts to that goal. Successful actions are additionally matched to standards [129] that are

19. Recent versions of SOAR (see [87]) incorporate many of ACT-R's features, including the distinction between procedural, declarative and episodic memory, and subsymbolic processes such as reinforcement learning and mental imagery.

associated with virtual goals to uphold them. If the appraisal of an action-relevant fact exceeds a threshold of intensity, up to three effects occur: The creation of an "impulse goal" to display a fitting emotional expression, the creation of a coping goal that may lead to a corresponding intention (e.g., retribution, trying to hinder), and a change of the evaluation of the agent who showed the action. These "emotion-enhanced" JAM agents have been used to steer the characters of a simple interactive drama featuring a hero, a victim, a mentor, and a villain.

A somewhat different approach to enhancing BDI-agents with emotions has been taken by Becker-Asano and Wachsmuth ([20]; see also [19]). In the emotional BDI agent *Max* constructed by these authors, an emotion system (a model of the pleasure-arousal theory of emotion; [162]) is added to the BDI architecture as a separate module that receives signals from, and sends signals to, several other modules. For example, on the input side, the emotion module receives a positive signal if a goal is achieved, which results in an increase in pleasure; whereas on the output side, emotions (pleasure-arousal states) among other effects generate beliefs about having them, amounting to self-awareness of the agent of its own emotions. In yet another "emotional" extension of the BDI architecture, Lorini and Piunti [97] have modified the processing cycle of the JADEX agent by including a surprise-based filter for belief revision ([97]; see also [100]).

To allow the modeling of different emotion theories or specific assumptions of different theories (see Section 2), it would be useful if prospective emotion modelers could more easily customize and extend BDI systems than is currently possible. One way of achieving this would be to break up the original BDI interpreter cycle into a small set of processes that can be combined as needed [137], as well as extended with additional processes. For example, it would be useful to be able to extend the BDI architecture with a set of basic desires from which concrete desires can be autonomously derived [11], [99], or with a more detailed model of memory similar to ACT-R (e.g., to model effects of emotions on memory). Also, the BDI architecture could be parameterized to allow the modeling of certain personality differences (e.g., [76], [125]). In fact, some of the proposed extensions are already being realized; for example, Ritter [159] has enriched a BDI agent architecture with elements from ACT-R in CoJACK.

4.3 Modeling Emotions in Affective Agent Architectures

Several specialized "affective" agent architectures have been developed by affective computing researchers. Some of these are (relatively) general-purpose architectures and thus suited to model different emotion theories. To illustrate this approach, we describe FAtiMA Modular [52]; [11] and MAMID [77], [79].

4.3.1 Modeling Emotions in FatiMA Modular

FAtiMA Modular is an autonomous agent architecture with reactive and deliberative (BDI-based) behavior designed to allow emotions and personality characteristics to influence the agent's behavior. It was created by refactoring a previous version of the architecture [53] into a modular version, consisting of a core that defines the general processing cycle of the agent, and several specialized components that are specific realizations of the generic agent functions (e.g., specific appraisal computation and action selection functions). The basic processing cycle of FAtiMA agents, as well as several details of the architecture, have been influenced by the EMA model [68]. The basic processing cycle comprises the following steps:

1. perception of events;
2. on this basis, memory is updated and
3. appraisal processes are triggered;
4. an affective state is generated;
5. actions are executed based on goal-based planning (a BDI-style reasoning process); and
6. actions are executed.

As concerns emotion generation, FAtiMA combines a reactive layer that generates emotions on the basis of a set of domain-specific rules with a deliberative layer that generates emotions via an appraisal of the agents' current plans (see Section 2.2.3). The core of the deliberative module is a continuous planner that was extended to include probability information about actions and to perform emotion-focused coping strategies. When an event is perceived, active plans are updated and prospect-based emotions (hope, fear, satisfaction) based on the agent's plans and goals are generated.

Different cognitive emotion theories (or components of such theories) can be implemented in FAtiMA modular by adding specific affect generation functions. Likewise, specific action selection functions can be implemented by adding specific behavior components. This philosophy is in agreement with the "modular" philosophy of theory systematization endorsed in Section 2. Recently, FatiMA Modular has been further extended to include, among others, an autobiographical memory for remembering emotionally significant events, a module for basic motives (adopted from Dörner's [55], PSI theory; see also [12]) from which concrete goals are derived, and provisions for modeling cultural differences [11].

4.3.2 Modeling Emotions in MAMID

Similar to FAtiMA, MAMID [77], [79] implements an extended sense-think/feel-act processing cycle. In this case, the involved processes are:

1. sensory preprocessing (translating incoming data into task-relevant cues);
2. attention (filtering incoming cues and selecting a subset for processing);
3. situation assessment (integrating individual cues into an overall situation assessment);
4. expectation generation (projecting the current situation onto possible future states);
5. emotion generation (deriving emotions on the basis of a variety of cues);
6. goal selection; and
7. action selection (selecting the best action for goal achievement).

The computations of each module are supported by long-term memories represented as belief nets or rules. The memories of MAMID agents can vary in type and structure to reflect differences in individual histories, sensitivities, and responsiveness.

In MAMID, emotional states influence processing in three basic ways, the first two of which correspond to those featured in Section 2.2.4: First, emotions influence the allocation of processing resources and thus determine indirectly whether a specific cue or situation is processed or a specific goal is selected. Second, emotions directly influence which goals and actions are selected. Third, emotions can change the speed and capacity of the different modules of the architecture. This effect is meant to model system-wide subsymbolic (neuromodulatory) effects of emotions. In addition, both sensory processing and attention are affected by global architectural parameters that determine the speed and capacity of each processing module, as well as the ordering of the constructs within each module's processing queue. Together, these parameters allow to model biases for each of the basic processes (attention, situation assessment, expectation generation, emotion generation, and goal selection)

A key objective in developing MAMID has been to provide flexibility in the modeling of assumptions about the generation and particularly the effects of emotions (in fact, the primary focus of MAMID is on the modeling of *emotion effects* on attention, perception and cognition, as such effects have been neglected in previous computational models). The methodology used to achieve this flexibility combines two ideas: 1) a high degree of parameterization of the processes and structures that comprise the different modules of the architecture, and 2) the modeler's interactive access to these parameters and their weights. For example, the valence (pleasure-displeasure quality) and intensity of emotions are computed in MAMID as linear combinations of several eliciting factors. The weights controlling the contributions of the individual eliciting factors can be modified interactively by the model developer. This provides a simple and quick way of implementing alternative hypotheses regarding the eliciting conditions of different emotions, as well as of tuning models to reflect emerging empirical data. Similarly, emotion-relevant individual differences (e.g., differences in threat sensitivity) can be modeled by interactively varying parameters. For example, in a stressful encounter, a trait-anxious agent is likely to consider its competence and degree of control, whereas a nontrait anxious agent would not consider these factors. This methodology for modeling the effects of emotions as well other behavior moderators (e.g., personality traits, see [158]) has several advantages [77], including the possibility to rapidly model a broad range of distinct individual profiles and to integrate the (possibly conflicting) effects of multiple interacting emotions and traits.

5 Conclusion

The proposals made in this paper for increasing the interdisciplinary exchange between psychology and computer science, as well as the intradisciplinary coordination, in the field of computational emotion modeling share two overarching themes. The first is a proposal for *modularization*: deconstruct emotion theories into basic assumptions; modularize architectures. The second is a proposal for *unification and standardization*: Translate different emotion theories into a common informal conceptual system or a formal language, or implement them in a common architecture. In making these proposals, we find ourselves in agreement with developments already underway in the domain of cognitive architectures. Specifically, Taatgen and Anderson [194] predict that, given that different cognitive architectures already share many identical or similar mechanisms, they will converge even more in the future, as a consequence of which the choice of a particular architecture will no longer imply strong theoretical commitments. Furthermore, by increasing the degree of modularization of architectures, it will be possible to build models out of components and thus avoid the constraints imposed by fixed architectures. Analogously, one may envision the development of a theoretical toolbox of basic theory-elements, formulated in a common language, from which theories of emotional agents (or of emotion modules for agents) can be constructed.

References

[1] C. Adam, A. Herzig, and D. Longin, "A Logical Formalization of the OCC Theory of Emotions," *Synthese,* vol. 168, pp. 201-248, 2009.

[2] I. Ajzen, "The Theory of Planned Behavior," *Organizational Behavior and Human Decision Processes,* vol. 50, pp. 179-211, 1991.

[3] S.R. Allen, "Concern Processing in Autonomous Agents," PhD thesis, Faculty of Science of the Univ. of Birmingham, School of Computer Science, 2001.

[4] A. Anderson, *The Architecture of Cognition.* Harvard Univ. Press, 1983.

[5] J.R. Anderson, D. Bothell, M.D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An Integrated Theory of the Mind," *Psychological Rev.,* vol. 111, pp. 1036-1060, 2004.

[6] *The Atomic Components of Thought,* J.R. Anderson and C. Lebiere, eds. Erlbaum, 1998.

[7] A.D. Angie, S. Connelly, E.O. Waples, and V. Kligyte, "The Influence of Discrete Emotions on Judgement and Decision-Making: A Meta-Analytic Review," *Cognition and Emotion,* vol. 25, pp. 1393-1422, 2011.

[8] J.L. Armony, D. Servan-Schreiber, J.D. Cohen, and J.E. LeDoux, "Computational Modeling of Emotion: Explorations through the Anatomy and Physiology of Fear Conditioning," *Trends in Cognitive Sciences,* vol. 1, pp. 28-34, 1997.

[9] M.B. Arnold, *Emotion and Personality.* Columbia Univ. Press, 1960.

[10] J.W. Atkinson, *An Introduction to Motivation.* Van Nostrand, 1964.

[11] R. Aylett and A. Paiva, "Computational Modelling of Culture and Affect" *Emotion Rev.,* vol. 4, pp. 253-263, 2012.

[12] J. Bach, *Principles of Synthetic Intelligence: An Architecture for Motivated Cognition.* Oxford Univ. Press, 2009.

[13] W. Balzer and C.U. Moulines, "Introduction," *Structuralist Knowledge Representations: Paradigmatic Examples (Poznan Studies in the Philosophy of the Sciences and the Humanities),* W. Balzer, J.D. Sneed, and C.U. Moulines, vol. 75, pp. 3-18, Rodopi, 2000.

[14] W. Balzer, C.U. Moulines, and J.D. Sneed, *An Architectonic for Science: The Structuralist Program.* Reidel, 1987.

[15] W. Balzer, J.D. Sneed, and C.U. Moulines, "Structuralist Knowledge Representations: Paradigmatic Examples," *Poznan Studies in the Philosophy of the Sciences and the Humanities,* vol. 75, Rodopi, 2000.

[16] L.F. Barrett, "Solving the Emotion Paradox: Categorization and the Experience of Emotion," *Personality and Social Psychology Rev.,* vol. 10, pp. 20-46, 2006.

[17] L.F. Barrett, "Are Emotions Natural Kinds?" *Perspectives on Psychological Science,* vol. 1, pp. 28-58, 2006.

[18] R.F. Baumeister, K.D. Vohs, C.N. DeWall, and L. Zhang, "How Emotion Shapes Behavior: Feedback, Anticipation, and Reflection, Rather Than Direct Causation," *Personality and Social Psychology Rev.,* vol. 11, pp. 167-203, 2007.

[19] C. Becker-Asano, *WASABI: Affect Simulation for Agents with Believable Interactivity.* IOS Press, 2008.

[20] C. Becker-Asano and I. Wachsmuth, "Affect Simulation with Primary and Secondary Emotions," *Intelligent Virtual Agents,* vol. 8, pp. 15-28, 2008.

[21] J. Bennett, "Analysis without Noise," *Mind and Common Sense,* R.J. Bogdan, ed., pp. 15-36, Cambridge Univ. Press, 1991.

[22] J. Bentham, *An Introduction to the Principles of Morals and Legislation.* Athlone Press, 1789/1970.

[23] I. Blanchette and A. Richards, "The Influence of Affect on Higher Level Cognition: A Review of Research on Interpretation, Judgement, Decision Making and Reasoning," *Cognition and Emotion,* vol. 24, pp. 561-595, 2010.

[24] N. Block, "Introduction: What Is Functionalism?" *Readings in Philosophy of Psychology,* N. Block, ed., vol. 1, pp. 171-184, Harvard Univ. Press, 1980.

[25] M. Boden, *Computer Models of Mind: Computational Approaches in Theoretical Psychology.* Cambridge Univ. Press, 1988.

[26] K. Bollen, *Structural Equations with Latent Variables.* Wiley, 1989.

[27] T. Bosse, J. Gratch, J.F. Hoorn, M. Pontier, and G.F. Siddiqui, "Comparing Three Computational Models of Affect," *Advances in Practical Applications of Agents and Multiagent Systems,* Y. Demazeau, F. Dignum, J. Corchado, and J. Prez, eds., vol. 70, pp. 175-184, Springer, 2010.

[28] N. Bourbaki, *Elements of Mathematics: Theory of Sets.* Addison-Wesley, 1968.

[29] M.E. Bratman, *Intentions, Plans, and Practical Reason.* Harvard Univ. Press, 1987.

[30] M. Bratman, D.J. Israel, and M.E. Pollack, "Plans and Resource-Bounded Practical Reasoning," *Computational Intelligence,* vol. 4, pp. 349-355, 1988.

[31] J. Broekens, "Modeling the Experience of Emotion," *Int'l J. Synthetic Emotions,* vol. 1, pp. 1-17, 2010.

[32] J. Broekens, D. DeGroot, and W.A. Kosters, "Formal Models of Appraisal: Theory, Specification, and Computational Model," *Cognitive Systems Research,* vol. 9, pp. 173-197, 2008.

[33] J.A. Brook, "Psychoanalysis and Commonsense Psychology," *Annual of Psychoanalysis,* Analytic Press, 1992.

[34] C.F. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton Univ. Press, 2003.

[35] W.B. Cannon, "The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory," *Am. J. Psychology,* vol. 39, pp. 106-124, 1927.

[36] C. Castelfranchi and E. Lorini, "Cognitive Anatomy and Functions of Expectations," *Proc. Int'l Joint Conf. Artificial Intelligence (IJCAI '03),* 2003.

[37] C. Castelfranchi and M. Miceli, "The Cognitive-Motivational Compound of Emotional Experience," *Emotion Rev.,* vol. 1, pp. 223-231, 2009.

[38] H-Q Chong, A-H Tan, and G-W Ng, "Integrated Cognitive Architectures: A Survey," *Artificial Intelligence Rev.,* vol. 28, pp. 103-130, 2007.

[39] R.E. Cochran, F.J. Lee, and E. Chown, "Modeling Emotion: Arousal's Impact on Memory," *Proc. 28th Ann. Conf. Cognitive Science Soc.,* pp. 1133-1138, 2006.

[40] P.R. Cohen and H.J. Levesque, "Intention Is Choice with Commitment," *Artificial Intelligence,* vol. 42, pp. 213-261, 1990.

[41] R. Conte and C. Castelfranchi, *Cognitive and Social Action.* Univ. College London Press, 1995.

[42] M.W. Cox and E. Klinger, "Motivation and the Theory of Current Concerns," *Handbook of Motivational Counseling,* W.M. Cox and E. Klinger, eds., pp. 1-28, Wiley, 2004.

[43] A.R. Damasio, *Descartes' Error.* Avon, 1994.

[44] M. Dastani, "2APL: A Practical Agent Programming Language," *Int'l J. Autonomous Agents and Multi-Agent Systems,* vol. 16, pp. 214-248, 2008.

[45] M. Dastani and J.-J.C. Meyer, "Agents with Emotions," *Int'l J. Intelligent Systems,* vol. 25, pp. 636-654, 2010.

[46] M. Dastani and E. Lorini, "A Logic of Emotions: From Appraisal to Coping," *Proc. 11th Int'l Conf. Autonomous Agents and Multiagent Systems (AAMAS '12),* pp. 1133-1140, 2012.

[47] M. Dastani, J. Hulstijn, and L. van der Torre, "How to Decide What to Do?" *European J. Operational Research,* vol. 160, pp. 762-784, 2005.

[48] W. Davis, "A Theory of Happiness," *Philosophical Studies,* vol. 39, pp. 305-317, 1981.

[49] R. Dawkins, "Universal Darwinism," *Evolution from Molecules to Man,* D.S. Bendall, ed., pp. 403-425, Cambridge Univ. Press, 1983.

[50] J.P. Day, "The Anatomy of Hope and Fear," *Mind,* vol. 79, pp. 369-384, 1970.

[51] D.C. Dennett, "Intentional Systems," *J. Philosophy,* vol. 68, pp. 87-106, 1971.

[52] J. Dias, S. Masceranhas, and A. Paiva, "Fatima Modular: Towards an Agent Architecture with a Generic Appraisal Framework," *Proc. Int'l Workshop Standards for Emotion Modeling,* 2011.

[53] J. Dias and A. Paiva, "Feeling and Reasoning: A Computational Model for Emotional Agents," *Proc. 12th Portuguese Conf. Artificial Intelligence (EPIA '05),* pp. 127-140, 2005.

[54] J. Dix and M. Fisher, "Where Logic and Agents Meet," *Annals of Math. and Artificial Intelligence,* vol. 61, pp. 15-28, 2011.

[55] D. Dörner, *Bauplan für eine Seele [Blueprint for a Mind].* Rowohlt, 2001.

[56] H.N. Duc, "Resource-Bounded Reasoning about Knowledge," PhD thesis, Faculty of Math. and Informatics, Univ. of Leipzig, 2001.

[57] R. Elio, "Belief-Desire-Intention Agency in a General Cognitive Architecture," *Cognitive Science Quarterly,* vol. 2, pp. 321-340, 2002.

[58] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi, *Reasoning About Knowledge.* MIT Press, 1995.

[59] *Expectations and Actions: Expectancy-Value Models in Psychology,* T.N. Feather, ed., Erlbaum, 1982.

[60] C.R. Fox and R.A. Poldrack, "Prospect Theory and the Brain," *Neuroeconomics: Decision Making and the Brain,* P.W. Glimcher, C.F. Camerer, E. Fehr, and R.A. Poldrack, eds., pp. 145-174, Elsevier, 2009.

[61] N.H. Frijda, "Problems of Computer Simulation," *Behavioral Science,* vol. 12, pp. 59-67, 1967.

[62] N.H. Frijda, *The Emotions.* Cambridge Univ. Press, 1986.

[63] N.H. Frijda, "Emotions Are Functional, Most of the Time," *The Nature of Emotion,* P. Ekman and R.J. Davidson, eds., pp. 112-136, Oxford Univ. Press, 1994.

[64] N.H. Frijda, A. Ortony, J. Sonnemans, and G.L. Clore, "The Complexity of Intensity: Issues Concerning the Structure of Emotion Intensity," *Rev. of Personality and Social Psychology,* M.S. Clark, ed., vol. 13, pp. 60-89, Sage, 1992.

[65] M. Georgeff, B. Pell, M. Pollack, M. Tambe, and M. Wooldridge, "The Belief-Desire-Intention Model of Agency," *Proc. Fifth Int'l Workshop Intelligent Agents: Agent Theories, Architectures, and Languages (ATAL '98),* pp. 1-10, 1999.

[66] J.H. Goldthorpe, "Rational Action Theory for Sociology," *British J. Sociology,* vol. 49, pp. 167-192, 1998.

[67] J. Gratch, "The Sciences of the Artificial Emotions: Comment on Aylett and Paiva," *Emotion Rev.,* vol. 4, pp. 266-268, 2012.

[68] J. Gratch and S. Marsella, "A Domain Independent Framework for Modeling Emotion," *J. Cognitive Systems Research,* vol. 5, pp. 269-306, 2004.

[69] J. Gratch, S. Marsella, N. Wang, and B. Stankovic, "Assessing the Validity of Appraisal-Based Models of Emotion," *Proc. Int'l Conf. Affective Computing and Intelligent Interaction,* 2009.

[70] O.H. Green, *The Emotions: A Philosophical Theory.* Kluwer, 1992.

[71] N. Guiraud, D. Longin, E. Lorini, S. Pesty, and J. Rivière, "The Face of Emotions: A Logical Formalization of Expressive Speech Acts," *Proc. Int'l Joint Conf. Autonomous Agents and Multiagent Systems (AAMAS '11),* pp. 1031-1038, 2011.

[72] E. Heise, "Anderson's ACT∗-Theory Applied to Skill Acquisition: from the Theoretical Framework to Empirical Content," *The Structuralist Program in Psychology: Foundations and Applications,* H. Westmeyer, ed., pp. 87-103, Hogrefe and Huber, 1992.

[73] K.V. Hindriks, "Programming Rational Agents in GOAL," *Multiagent Programming: Languages, Tools and Applications,* R.H. Bordini, M. Dastani, J. Dix, and A. El Fallah Seghrouchni, eds., pp. 119-157, Springer, 2009.

[74] K.V. Hindriks and J. Broekens, *Comparing Formal Cognitive Emotion Theories.* Delft Univ. of Technology, 2011.

[75] M.J. Huber, "JAM: A BDI-Theoretic Mobile Agent Architecture," *Proc. Third Int'l Conf. Autonomous Agents (Agents '99),* pp. 236-243, 1999.

[76] E. Hudlicka, "Modeling Emotion in Symbolic Cognitive Architectures," *Proc. AAAI Fall Symp.,* 1998.

[77] E. Hudlicka, "Two Sides of Appraisal: Implementing Appraisal and Its Consequences within a Cognitive Architecture," *Proc. AAAI Spring Symp., Architectures for Modeling Emotion,* 2004.

[78] E. Hudlicka, "Review of Cognitive-Affective Architectures," *Organizational Modeling: From Individuals to Societies,* G. Zacharias, J. McMillan, and S. Van Hemel, eds., pp. 167-184, Nat'l Academies Press, 2008.

[79] E. Hudlicka, "Modeling the Mechanisms of Emotion Effects on Cognition," *Proc. AAAI Fall Symp.: Biologically Inspired Cognitive Architectures,* pp. 82-86, 2008.

[80] E. Hudlicka, "Guidelines for Developing Computational Models of Emotions," *Int'l J. Synthetic Emotions,* vol. 2, pp. 26-79, 2011.

[81] G.E. Hughes and M.J. Cresswell, *A New Introduction to Modal Logic.* Routledge, 1996.

[82] W. James, *Principles of Psychology,* vols. 1-2. Dover, 1890/1950.

[83] R.M. Jones, C. Lebiere, and J.A. Crossman, "Comparing Modeling Idioms in ACT-R and Soar," *Proc. Eighth Int'l Conf. Cognitive Modeling (ICCM '07),* pp. 49-54, 2007.

[84] D. Kahneman and A. Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica,* vol. 47, pp. 263-291, 1979.

[85] D. Keltner and A.M. Kring, "Emotion, Social Function, and Psychopathology," *Rev. of General Psychology,* vol. 3, pp. 320-342, 1998.

[86] J.D. Laird, *Feelings: The Perception of Self.* Oxford Univ. Press, 2007.

[87] J.E. Laird, *The Soar Cognitive Architecture.* MIT Press, 2012.

[88] J.E. Laird, A. Newell, and P.S. Rosenbloom, "Soar: An Architecture for General Intelligence," *Artificial Intelligence,* vol. 33, pp. 1-64, 1987.

[89] P. Langley, "Cognitive Architectures and General Intelligent Systems," *AI Magazine,* vol. 27, pp. 33-44, 2006.

[90] P. Langley, J.E. Laird, and S. Rogers, "Cognitive Architectures: Research Issues and challenges," *Cognitive Systems Research,* vol. 10, pp. 141-160, 2009.

[91] R.S. Lazarus, "Thoughts on the Relations between Emotion and Cognition," *Am. Psychologist,* vol. 37, pp. 1019-1024, 1982.

[92] R.S. Lazarus, *Emotion and Adaptation.* Oxford Univ. Press, 1991.

[93] H. Leventhal and K.R. Scherer, "The Relationship of Emotion and Cognition: A Functional Approach to a Semantic Controversy," *Cognition and Emotion,* vol. 1, pp. 3-28, 1987.

[94] M. Lewis, J.M. Haviland-Jones, and L.F. Barrett, *Handbook of Emotions,* third ed. Guilford Press, 2008.

[95] E. Lorini, "A Dynamic Logic of Knowledge, Graded Beliefs and Graded Goals and Its Application to Emotion Modelling," *LORI-III Workshop Logic, Rationality and Interaction,* H. van Ditmarsch, J. Lang, and S. Ju, eds., pp. 165-178, 2011.

[96] E. Lorini, "The Cognitive Anatomy and Functions of Expectations Revisited," *The Goals of Cognition: Essays in Honor of Cristiano Castelfranchi,* F. Paglieri, L. Tummolini, R. Falcone, and M. Miceli, eds., pp. 273-288, College Publications, 2012.

[97] E. Lorini and M. Piunti, "The Benefits of Surprise in Dynamic Environments: From Theory to Practice," *Proc. Second Int'l Conf. Affective Computing and Intelligent Interaction,* pp. 362-373, 2007.

[98] E. Lorini and F. Schwarzentruber, "A Logic for Reasoning about Counterfactual Emotions," *Artificial Intelligence,* vol. 175, pp. 814-847, 2011.

[99] L. Macedo and A. Cardoso, "The Exploration of Unknown Environments Populated with Entities by a Surprise-Curiosity-Based Agent," *Cognitive Systems Research,* vol. 19-20, pp. 62-87, 2012.

[100] L. Macedo, A. Cardoso, R. Reisenzein, L. Lorini, and C. Castelfranchi, "Artificial Surprise," *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence,* J. Vallverdú, and D. Casacuberta, eds., IGI Global, 2009.

[101] G. Mandler, *Mind and Body.* Norton, 1984.

[102] K. Manhart, "Können AI-Programme als Theorien Betrachtet Werden? [Can AI Programs Be Regarded as Theories?]," *Proc. Fifth Austrian Meeting for Artificial Intelligence Conf.,* pp. 346-358, 1989.

[103] J.N. Marewski and K. Mehlhorn, "Using the ACT-R Architecture to Specify 39 Quantitative Process Models of Decision Making," *Judgment and Decision Making,* vol. 6, pp. 439-519, 2011.

[104] R. Marinier, J. Laird, and R. Lewis, "A Computational Unification of Cognitive Behavior and Emotion," *J. Cognitive Systems Research,* vol. 10, pp. 48-69, 2009.

[105] J. Marks, "A Theory of Emotion," *Philosophical Studies,* vol. 42, pp. 227-242, 1982.

[106] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* Freeman, 1982.

[107] S. Marsella and J. Gratch, "EMA: A Process Model Of Appraisal Dynamics," *J. Cognitive Systems Research,* vol. 10, pp. 70-90, 2009.

[108] S. Marsella, J. Gratch, and P. Petta, "Computational Models of Emotion," *Blueprint for Affective Computing: A Sourcebook,* K.R. Scherer, T. Bänziger, and E. Roesch, eds., pp. 21-46, Oxford Univ. Press, 2010.

[109] W. McDougall, *An Introduction to Social Psychology.* Methuen, 1908/1960.

[110] W. McDougall, *An Outline of Psychology.* Methuen, 1923.

[111] A. Meinong, *Psychologisch-Ethische Untersuchungen zur Werttheorie.* Leuschner and Lubensky, 1894.

[112] A.R. Mele, *Springs of Action: Understanding Intentional Behavior.* Oxford Univ. Press, 1992.

[113] B.A. Mellers, "Choice and the Relative Pleasure of Consequences," *Psychological Bull.,* vol. 126, pp. 910-924, 2000.

[114] J.-J.C. Meyer, "Reasoning about Emotional Agents," *Int'l J. Intelligent Systems,* vol. 21, pp. 601-619, 2006.

[115] J.-J.C. Meyer and W. van der Hoek, "Epistemic Logic for AI and Computer Science," *Cambridge Tracts in Theoretical Computer Science,* Cambridge Univ. Press, 1995.

[116] G. Miller, E. Galanter, and K.H. Pribram, *Plans and the Structure of Behavior.* Holt, 1960.

[117] S. Mitchell, "Function, Fitness and Disposition," *Biology and Philosophy,* vol. 10, pp. 39-54, 1995.

[118] D. Moffat, N.H. Frijda, and H. Phaf, "Analysis of a Model of Emotions," *Prospects for Artificial Intelligence,* A. Sloman, D. Hogg, G. Humphreys, A. Ramsay, and D. Partridge, eds., IOS Press, 1993.

[119] U. Neisser, *Cognitive Psychology.* Appleton-Century-Crofts, 1967.

[120] R.M. Nesse, "The Smoke Detector Principle: Natural Selection and the Regulation of Defenses," *Annals of the New York Academy of Sciences,* vol. 935, pp. 75-85, 2001.

[121] A. Newell, "You Can't Play 20 Questions with Nature and Win: Projective Comments on the Papers of This Symposium," *Visual Information Processing,* W.G. Chase, ed., pp. 283-308, Academic Press, 1973.

[122] A. Newell, "The Knowledge Level," *Artificial Intelligence,* vol. 18, pp. 87-127, 1982.

[123] A. Newell, *Unified Theories of Cognition.* Harvard, 1990.

[124] A. Newell and H.A. Simon, "GPS, a Program That Simulates Human Thought," *Computers and Thought,* E.A. Feigenbaum and J. Feldman, eds., pp. 279-293, McGraw-Hill, 1963.

[125] E. Norling and E. Ritter, "Towards Supporting Psychologically Plausible Variability in Agent-Based Human Modelling," *Proc. Third Int'l Joint Conf. Autonomous Agents and Multiagent Systems (AAMAS '04),* pp. 758-765, 2004.

[126] K. Oatley, "Communications to Self and Others: Emotional Experience and Its Skills," *Emotion Rev.,* vol. 1, pp. 206-213, 2009.

[127] K. Oatley and P.N. Johnson-Laird, "Towards a Cognitive Theory of Emotions," *Cognition and Emotion,* vol. 1, pp. 29-50, 1987.

[128] A. Öhman and S. Mineka, "Fears, Phobias, and Preparedness: Towards an Evolved Module of Fear and Fear Learning," *Psychological Rev.,* vol. 108, pp. 483-522, 2001.

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

[130] A. Ortony, D. Norman, and W. Revelle, "Affect and Proto-Affect in Effective Functioning," *Who Needs Emotions: The Brain Meets the Machine,* J.M. Fellous and M.A. Arbib, eds., Oxford Univ. Press, 2005.

[131] S.E. Palmer, "Fundamental Aspects of Cognitive Representation," *Cognition and Categorization,* E. Rosch and B.B. Lloyd, eds., pp. 259-302, Erlbaum, 1978.

[132] N. Peira, A. Golkar, A. Öhman, S. Anders, and S. Wiens, "Emotional Responses in Spider Fear Are Closely Related to Picture Awareness," *Cognition and Emotion,* vol. 26, pp. 252-260, 2012.

[133] D. Pereira, E. Oliveira, and N. Moreira, "Formal Modelling of Emotions in BDI Agents," *Proc. Eighth Workshop Computational Logic in Multi-Agent Systems,* pp. 62-81, 2008.

[134] R. Pfeifer, "Symbols, Patterns, and Behavior: Towards a New Understanding of Intelligence," *Proc. Conf. Artificial Intelligence,* pp. 1-15, 1996.

[135] R.W. Picard, *Affective Computing.* MIT Press, 1997.

[136] A. Pokahr, L. Braubach, and W. Lamersdorf, "JADEX: A BDI Reasoning Engine," *Multi-Agent Programming: Languages, Platforms and Applications,* R.H. Bordini, M. Dastani, J. Dix and A. El Fallah-Seghrouchni, eds., pp. 149-174, Springer, 2005.

[137] A. Pokahr, W. Braubach, and W. Lamersdorf, "A Flexible BDI Architecture Supporting Extensibility," *Proc. IEEE/WIC/ACM Int'l Conf. Intelligent Agent Technology,* pp. 379-385, 2005b.
[138] J.L. Pollock, "OSCAR: A General Theory of Rationality," *J. Experimental and Theoretical Artificial Intelligence,* vol. 1, pp. 209-226, 1989.
[139] D.V. Pynadath and S. Marsella, "PsychSim: Modeling Theory of Mind with Decision-Theoretic Agents," *Proc. Int'l Joint Conf. Artificial Intelligence,* pp. 1181-1186, 2005.
[140] S. Rank, "Affective Acting: An Appraisal-Based Architecture for Agents as Actors," Master's thesis, Institut für Medizinische Kybernetik und Artificial Intelligence, Univ. of Vienna, http://www.ofai.at/~stefan.rank/StefanRank-AAAThesis.pdf, 2004.
[141] S. Rank, "Towards Reusable Roleplayers Using an Appraisal-Based Architecture," *Applied Artificial Intelligence,* vol. 19, pp. 313-340, 2005.
[142] A.S. Rao and M.P. Georgeff, "Modeling Rational Agents within a BDI-Architecture," *Proc. Second Int'l Conf. Principles of Knowledge Representation and Reasoning (KR '91),* pp. 473-484, 1991.
[143] A.S. Rao and M.P. Georgeff, "BDI Agents: From Theory to Practice," *Proc. First Int'l Conf. Multiagent Systems (ICMAS '95),* pp. 312-319, 1995.
[144] R. Reisenzein, "Pleasure-Arousal Theory and the Intensity of Emotions," *J. Personality and Social Psychology,* vol. 67, pp. 525-539, 1994.
[145] R. Reisenzein, "Emotional Action Generation" *Processes of the Molar Regulation of Behavior,* W. Battmann and S. Dutke, eds., pp. 151-165, Pabst Science Publishers, 1996.
[146] R. Reisenzein, "Wundt's Three-Dimensional Theory of Emotion," *Structuralist Knowledge Representations: Paradigmatic Examples (Poznan Studies in the Philosophy of the Sciences and the Humanities),* W. Balzer, J.D. Sneed, and C.U. Moulines, eds., vol. 75, pp. 219-250, Rodopi, 2000.
[147] R. Reisenzein, "Appraisal Processes Conceptualized from a Schema-Theoretic Perspective: Contributions to a Process Analysis of Emotions," *Appraisal Processes in Emotion: Theory, Methods, Research,* K.R. Scherer, A. Schorr, and T. Johnstone, eds., pp. 187-201, Oxford Univ. Press, 2001.
[148] R. Reisenzein, "Arnold's Theory of Emotion in Historical Perspective," *Cognition and Emotion,* vol. 20, pp. 920-951, 2006.
[149] R. Reisenzein, "What Is a Definition of Emotion? And Are Emotions Mental-Behavioral Processes?" *Social Science Information,* vol. 46, pp. 424-428, 2007.
[150] R. Reisenzein, "Emotional Experience in the Computational Belief-Desire Theory of Emotion," *Emotion Rev.,* vol. 1, pp. 214-222, 2009.
[151] R. Reisenzein, "Emotions as Metarepresentational States of Mind: Naturalizing the Belief-Desire Theory of Emotion," *Cognitive Systems Research,* vol. 10, pp. 6-20, 2009.
[152] R. Reisenzein, "Broadening the Scope of Affect Detection Research," *IEEE Trans. Affective Computing,* vol. 1, no. 1, pp. 42-45, Jan.-June 2010.
[153] R. Reisenzein, "What Is an Emotion in the Belief-Desire Theory of Emotion?" *The Goals of Cognition: Essays in Honor of Cristiano Castelfranchi,* F. Paglieri, L. Tummolini, R. Falcone, and M. Miceli, eds., pp. 181-211, College Publications, 2012.
[154] R. Reisenzein and S. Döring, "Ten Perspectives on Emotional Experience: Introduction to the Special Issue," *Emotion Rev.,* vol. 1, pp. 195-205, 2009.
[155] R. Reisenzein and G. Horstmann, *Emotion,* third ed., H. Spada, ed., pp. 435-500. Huber, 2006.
[156] R. Reisenzein and M. Junge, "Überraschung, Enttäuschung und Erleichterung: Emotionsintensität als Funktion von subjektiver Wahrscheinlichkeit und Erwünschtheit [Surprise, Disappointment and Relief: Emotion Intensity as Function of Subjective Probability and Desirability]," *Proc. 45th Congress German Psychological Assoc. in Nuremburg,* 2006.
[157] R. Reisenzein and M. Junge, "Language and Emotion from the Perspective of the Computational Belief-Desire Theory of Emotion," *Dynamicity in Emotion Concepts,* P.A. Wilson, ed., vol. 27, pp. 37-59. Peter Lang, 2012.
[158] R. Reisenzein and H. Weber, "Personality and Emotion," *Cambridge Handbook of Personality Psychology,* P.J. Corr and G. Matthews, eds., pp. 54-71, Oxford Univ. Press, 2009.
[159] F.E. Ritter, "CoJACK: A High-Level Cognitive Architecture with Demonstrations of Moderators, Variability, and Implications for Situation Awareness," *Biologically Inspired Cognitive Architectures,* vol. 1, pp. 2-13, 2012.
[160] F.E. Ritter, M. Avraamides, and I.G. Councill, "Validating Changes to a Cognitive Architecture to More Accurately Model the Effects of Two Example Behavior Moderators," *Proc. 11th Computer-Generated Forces and Behavior Representation Conf.,* pp. 29-40, 2002.
[161] I.J. Roseman, "A Model of Appraisal in the Emotion System." *Appraisal Processes in Emotion: Theory, Methods, Research,* K.R. Scherer, A. Schorr, and T. Johnstone, eds., pp. 68-91, Oxford Univ. Press, 2001.
[162] J.A. Russell, "A Circumplex Model of Affect," *J. Personality and Social Psychology,* vol. 39, pp. 1161-1178, 1980.
[163] J.A. Russell, "Core Affect and the Psychological Construction of Emotion," *Psychological Rev.,* vol. 110, pp. 145-172, 2003.
[164] S.J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach,* third ed. Prentice Hall, 2010.
[165] S. Schachter, "The Interaction of Cognitive and Physiological Determinants of Emotional State," *Advances in Experimental Social Psychology,* L. Berkowitz, ed., vol. 1, pp. 49-80, Academic Press, 1964.
[166] K.R. Scherer, "Appraisal Considered as a Process of Multilevel Sequential Checking," *Appraisal Processes in Emotion: Theory, Methods, Research,* K.R. Scherer, A. Schorr, and T. Johnstone, eds., pp. 92-129, Oxford Univ. Press, 2001.
[167] K.R. Scherer, "Emotion and Emotional Competence: Conceptual and Theoretical Issues for Modelling Agents," *Blueprint for Affective Computing: A Sourcebook,* K.R. Scherer, T. Bänziger, and E. Roesch, eds., pp. 3-20, Oxford Univ. Press, 2010.
[168] K.R. Scherer, T. Bänziger, and E. Roesch, *Blueprint for Affective Computing: A Sourcebook.* Oxford Univ. Press, 2010.
[169] M. Schut and M. Wooldridge, "The Control of Reasoning in Resource-Bounded Agents," *Knowledge Eng. Rev.,* vol. 16, pp. 215-240, 2001.
[170] N. Schwarz and G.L. Clore, "Feelings and Phenomenal Experiences," *Social Psychology: Handbook of Basic Principles,* second ed., E.T. Higgins and A.W. Kruglanski, eds., pp. 385-407, Guilford, 2007.
[171] J. Searle, *Intentionality.* Cambridge Univ. Press, 1983.
[172] J.Y. Shah and W.L. Gardner, *Handbook of Motivation Science.* Guildford Press, 2008.
[173] M. Siemer and R. Reisenzein, "The Process of Emotion Inference," *Emotion,* vol. 7, pp. 1-20, 2007.
[174] M. Siemer and R. Reisenzein, "Appraisals and Emotions: Can You Have One without the Other?" *Emotion,* vol. 7, pp. 26-29, 2007.
[175] M. Siemer, I. Mauss, and J.J. Gross, "Same Situation-Different Emotions: How Appraisals Shape Our Emotions," *Emotion,* vol. 7, pp. 592-600, 2007.
[176] H.A. Simon, "Motivational and Emotional Controls of Cognition," *Psychological Rev.,* vol. 74, pp. 29-39, 1967.
[177] B.D. Slife and R.N. Williams, "Toward a Theoretical Psychology: Should a Subdiscipline Be Formally Recognized?" *Am. Psychologist,* vol. 52, pp. 117-129, 1997.
[178] A. Sloman, "Prolegomena to a Theory of Communication and Affect," *Comm. from an Artificial Intelligence Perspective: Theoretical and Applied Issues,* A. Ortony, J. Slack, and O. Stock, eds., pp. 229-260, Springer, 1992.
[179] A. Sloman and M. Croucher, "Why Robots Will Have Emotions," *Proc. Seventh Int'l Joint Conf. Artificial Intelligence,* pp. 197-202, 1981.
[180] P. Slovic, M. Finucane, E. Peters, and D. MacGregor, "Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality," *Risk Analysis,* vol. 24, pp. 311-322, 2004.
[181] C.A. Smith and L.D. Kirby, "Consequents Require Antecedents: Toward a Process Model of Emotion Elicitation," *Feeling and Thinking: The Role of Affect in Social Cognition,* J.P. Forgas, ed., pp. 83-105, Cambridge Univ. Press, 2000.
[182] Smith, C.A., and R.S. Lazarus, "Emotion and Adaptation," *Handbook of Personality: Theory and Research,* L. Pervin, ed., pp. 609-637, Guilford, 1990.
[183] M. Smith, *The Moral Problem.* Blackwell, 1997.
[184] W. Stegmüller, *Theorie und Erfahrung. Dritter Teilband: Die Entwicklung des neueren Strukturalismus seit 1973. [Theory and Experience, Vol. 3, [The Development of Recent Structuralism Since 1973].* Springer, 1986.
[185] B.R. Steunebrink, M. Dastani, and J.-J.C. Meyer, "A Formal Model of Emotions: Integrating Qualitative and Quantitative Aspects," *Proc. 18th European Conf. Artificial Intelligence* pp. 256-260, 1986.

[186] B.R. Steunebrink, M. Dastani, and J.-J.C. Meyer, "The OCC Model Revisited," *Proc. Fourth Workshop Emotion and Computing—Current Research and Future Impact,* pp. 40-47, 1986.
[187] B.R. Steunebrink, M. Dastani, and J.-J.C. Meyer, "A Formal Model of Emotion Triggers: An Approach for BDI Agents," *Synthese,* vol. 185, pp. 83-129, 2012.
[188] A. Stocco, D. Fum, and T. Zalla, "Revising the Role of Somatic Markers in the Gambling Task: A Computational Account for Neuropsychological Impairments," *Proc. 27th Ann. Conf. Cognitive Science Soc.,* 2005.
[189] J. Storbeck and G.L. Clore, "Emotional Controls on Cognitive Processes," *Cognition and Emotion,* vol. 21, pp. 1212-1237, 2007.
[190] K.T. Strongman, *The Psychology of Emotion: From Everyday Life to Theory,* fifth ed. Wiley, 2003.
[191] R. Sun, *Cambridge Handbook of Computational Psychology.* Cambridge Univ. Press, 2008.
[192] R. Sun, "Introduction to Computational Cognitive Modeling," *Cambridge Handbook of Computational Psychology,* R. Sun, ed., pp. 3-19, Cambridge Univ. Press, 2008.
[193] N.A. Taatgen and J.R. Anderson, "Constraints in Cognitive Architectures," *Cambridge Handbook of Computational Psychology,* R. Sun, ed., pp. 170-185, Cambridge Univ. Press, 2008.
[194] N.A. Taatgen and J.R. Anderson, "The Past, Present, and Future of Cognitive Architectures," *Topics in Cognitive Science,* vol. 4, pp. 693-704, 2010.
[195] C.C.W. Taylor, "Emotions and Wants," *The Ways of Desire,* J. Marks, ed., pp. 217-231, Precedent, 1986.
[196] J. Tooby and L. Cosmides, "The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environments," *Ethology and Sociobiology,* vol. 11, pp. 375-424, 1990.
[197] K.G. Troitzsch, "Simulation and Structuralism," *Modelling and Simulation in the Social Sciences from a Philosophy of Science Point of View,* R. Hegselmann, U. Mueller, and K.G. Troitzsch, eds., pp. 183-208, Kluwer, 1996.
[198] K.G. Troitzsch, "Theory Reconstruction of Several Versions of Modern Organisation Theories," *Ontology, Epistemology, and Teleology for Modeling and Simulation: Philosophical Foundations for Intelligent M&S Applications,* A. Tolk, ed., pp. 121-140, Springer, 2012.
[199] W. van der Hoek and M. Wooldridge, "Towards a Logic of Rational Agency," *Logic J. GPL,* vol. 11, pp. 133-157, 2003.
[200] B. van Linder, W. van der Hoek, and J.-J.C. Meyer, "Formalizing Abilities and Opportunities of Agents," *Fundamenta Informaticae,* vol. 34, pp. 53-101, 1998.
[201] D. Vernon, G. Metta, and G. Sandini, "A Survey of Artificial Cognitive Systems: Implications for the Autonomous Development of Mental Capabilities in Computational Agents," *IEEE Trans. Evolutionary Computation,* vol. 11, no. 2, pp. 151-180, Apr. 2007.
[202] B. Weiner, *Human Motivation: Metaphors, Theory and Research.* Wiley, 1992.
[203] B. Weiner, *Judgments of Responsibility: A Foundation for a Theory of Social Conduct.* Guilford, 1995.
[204] *The Structuralist Program in Psychology: Foundations and Applications,* H. Westmeyer, ed. Hogrefe & Huber, 1992.
[205] H. Westmeyer, *Introduction to the Structuralist Program in Psychology,* H. Westmeyer, ed., pp. 1-12. Hogrefe & Huber, 1992.
[206] M.J. Wooldridge, *Reasoning about Rational Agents.* MIT Press, 2000.
[207] W. Wundt, *Grundriss der Psychologie.* Engelmann, 1896.
[208] G.L. Zacharias, J. Macmillan, and B.S. Van Hemel, *Behavioral Modeling and Simulation: From Individuals to Societies.* Nat'l Academies Press, 2008.
[209] R.B. Zajonc, "Feeling and Thinking: Preferences Need No Inferences," *Am. Psychologist,* vol. 35, pp. 151-175, 1980.

Rainer Reisenzein received the PhD degree in psychology from the University of Salzburg, Austria. After spending a postdoctoral research year at the University of California, Los Angeles, he was an assistant professor of psychology at the Free University of Berlin and then at the University of Bielefeld. Since 2002, he has been a full professor of general psychology at the University of Greifswald, Germany. His research interests include theoretical and empirical questions related to emotion and motivation and is marked by an interdisciplinary orientation toward, in particular, philosophy (philosophy of science and philosophy of mind) and cognitive science. He has published more than 100 theoretical, review, and empirical articles and book chapters in English and German on diverse topics, mostly related to emotion, including emotional experience, emotion concepts, the facial expression of emotion, basic emotions, cognitive emotion theory, and the history of emotion research. A special research focus is the phenomenon of surprise in humans, considered from the perspective of emotion theory. He is also a coauthor (with Wulf-Uwe Meyer und Achim Schützwohl) of a German three-volume introduction to the psychology of emotions. In 2009, he edited (together with philosopher Sabine Döring) a special issue of *Emotion Review* on the topic of emotional experience. He is a member of the HUMAINE Association and of the International Society for Research on Emotion, and he is currently on the editorial board of the journals *Cognition and Emotion* and *Experimental Psychology.*

Eva Hudlicka received the BS degree in biochemistry from Virginia Tech, the MS degree in computer science from the Ohio State University, and the PhD degree in computer science from the University of Massachusetts-Amherst. Subsequently, she was a senior scientist at Bolt Beranek and Newman, Cambridge, Massachusetts. In 1995, she founded Psychometrix Associates, Inc. to pursue research in affective computing. Her primary research interests include the development of computational models of emotion, both the cognitive processes involved in appraisal, and the effects of emotions on cognition. The aim of this work is an improved understanding of the mechanisms of emotion generation and emotion-cognition interactions, including the nature of affective biases on cognition. The findings have practical implications for the development of affective user models and affective virtual agents across a range of applications, including serious games for training and therapy, and for cognitive-affective engineering in general. Specific applications include training and education, decision-support systems, assistive technologies, assessment and treatment of cognitive-affective disorders, and affective gaming. Her research is conducted within the context of a computational cognitive-affective architecture developed by Psychometrix, the MAMID architecture (Methodology for Modeling and Analysis of Individual Differences), which implements a generic methodology for modeling the interacting effects of multiple affective factors and individual differences on decision making. In 2008, she was a member of a National Research Council Committee on Organizational Models: From Individuals to Societies, contributing her expertise on modeling individual behavior with a focus on affective processes. She is an associate editor of the *International Journal of Synthetic Emotions* and a member of the editorial board of the Oxford University Press series on *Cognitive Models and Architectures*, the *International Journal of Machine Consciousness*, and the *International Journal of Biologically Inspired Cognitive Architectures*. She has taught courses and tutorials on affective computing, affective computing for game design, and computational emotion modeling, and has authored numerous book chapters, articles, and reports in this research area.

Mehdi Dastani received the PhD degree in humanities from the University of Amsterdam for his dissertation titled "Languages of Perception," and the MS degree in computer science and the MS degree in philosophy from the University of Amsterdam. He is an associate professor of computer science at Utrecht University in the Netherlands. His research interests include intelligent systems, in particular logical and computational models of social and cognitive phenomena such as emotion, norms, decision making, interaction, and perception. Currently, he is working in the area of multiagent systems, multiagent logics, and multiagent programming. He has developed various multiagent (programming) frameworks such as BOID (Beliefs, Obligations, Intentions and Desires), 2APL (A Practical Agent Programming Language), and 2OPL (Organization-Oriented Programming Language), which support the development and implementation of multiagent systems. These frameworks are being used by many universities for teaching and training purposes. An important line of research in his work concerns emotion modeling. In this domain, he has developed several logical models for emotion processes such as appraisal, emotion experience and coping, which are formalizations of the psychological emotion models proposed by Ortony, Clore and Collins, Frijda, and Lazarus. In addition to being the author of numerous technical articles, he has co-organized numerous international events related to multiagent systems, including various editions of the International conference of Autonomous Agents and Multi-Agent Systems (AAMAS), the International workshop Multi-Agent Programming (ProMAS), the Lorentz seminar on Multi-Agent Organisations and the Dagstuhl seminar on Multi-Agent Programming. He was also on the chair of the European Agent System Summer Schools (EASSS) between 2007 and 2010, the 19th Belgian-Dutch Conference on Artificial Intelligence (BNAIC 2007), and the Fifth European Workshop Multi-Agent Systems (EUMAS 2007). In addition, he has been a member of the senior programme committee of many international conferences on artificial intelligence (e.g., IJCAI, ECAI, AAAI, AAMAS).

Jonathan Gratch received the PhD degree in computer science from the University of Illinois in Urban-Champaign in 1995. He is an associate director of Virtual Human Research at the University of Southern California's (USC) Institute for Creative Technologies, research associate professor of computer science and psychology at USC, and codirector of USC's Computational Emotion Group. His research interests include computational models of human cognitive and social processes, especially emotion, and explores these models' role in shaping human-computer interactions in virtual environments. He studies the relationship between cognition and emotion, the cognitive processes underlying emotional responses, and the influence of emotion on decision making and physical behavior. He is the author of more than 200 technical articles. He is the founding and current editor-in-chief of the *IEEE Transactions on Affective Computing*, associate editor of *Emotion Review* and the *Journal of Autonomous Agents and Multiagent Systems*, former president of the HUMAINE Association, and is a member of the IEEE, the Association for the Advancement of Artificial Intelligence, and the International Society for Research on Emotion.

Koen Hindriks received the PhD degree at Utrecht University on agent programming languages. He is an assistant professor with the Interactive Intelligence Group with the Faculty of Electrical Engineering, Mathematics, and Computer Science at the Delft University of Technology, The Netherlands. His main research interests include cognitive agent technology and coordination models for effective multiagent interaction. His research focuses on the analysis, modeling, and development of agent technology that integrates different aspects of intelligence such as reasoning, decision making, planning, learning and interaction, as well as aspects of emotional intelligence. This multiagent technology has been applied, among others, in microsimulations of domains such as traffic, logistics and supply chain management, serious gaming, negotiation, sociocognitive robotics, and user modeling. He has designed and developed the agent programming languages 3APL and GOAL and has worked on the verification and specification of agent programs. He has published more than 100 articles on agent technology and organized various events in this area, including the 2012 Dagstuhl Seminar Engineering Multi-Agent Systems.

Emiliano Lorini is a research scientist at the Institut de Recherche en Informatique de Toulouse (IRIT) of the French National Research Council (CNRS). His research topics include epistemic logic and formal epistemology, game theory, logics of intention, logics of agency, deontic logic, the theory of emotion, and the theory of trust and reputation. He has authored or coauthored more than 80 articles in journals and international conferences in the fields of AI, logic, cognitive science, and game theory. He has been a reviewer for major journals and conferences in the areas of AI, cognitive science, and economics, including *AI Journal*, *Mathematical Social Sciences*, *Journal of Autonomous Agents and Multi-Agent Systems*, *Synthese*, *Cognitive Systems Research*, *Journal of Philosophical Logic*, AAAI, IJCAI, KR, AAMAS, and ECAI. He has been a member of the program committee of several major conferences in the field of AI and cognitive science, including AAMAS, ECAI, IJCAI 2011, and CogSci, and has co-organized several conferences and workshops in the fields of AI, game theory, and logic. He has also been editor of special issues of, respectively, the *Journal of Philosophical Logic* and *Synthese*.

John-Jules Ch. Meyer received the PhD degree in theoretical computer science from the VU University in Amsterdam. He studied mathematics with computer science and digital signal processing at Leyden University. From 1988 to 1993, he was a full professor both at the VU University in Amsterdam and the University of Nijmegen. Since 1993, he has been a full professor of computer science at Utrecht University (UU). Currently, he is heading the Artificial Intelligence Section within the Department of Information and Computing Sciences and is program leader of the AI Master of the UU. He is also the CSO of the Alan Turing Institute Almere, which specializes in multidisciplinary research for the health sciences. He has more than 400 peer-reviewed publications in international journals and conference proceedings, and an h-index of 42 (Google scholar). He is a member of the editorial boards of several international journals, including the *Journal of Intelligent Agents and Multi-Agent Systems*. His current research interests include artificial intelligence, and intelligent agents in particular, as to both theory (e.g., agent logics) and applications such as virtual characters in serious games. In 2005, he was appointed as a fellow of the European Coordinating Committee for Artificial Intelligence. He has been an honorary member of the School of Information and Knowledge Systems since 2007.