Available online at www.sciencedirect.com
ScienceDirect
Cognitive Systems Research 10 (2009) 48–69
Cognitive Systems RESEARCH
www.elsevier.com/locate/cogsys

A computational unification of cognitive behavior and emotion

Action editor: Jonathan Gratch

Robert P. Marinier III [a,*], John E. Laird [a], Richard L. Lewis [b]

[a] University of Michigan, Department of Electrical Engineering and Computer Science, 2260 Hayward Street, Ann Arbor, MI 48109, USA
[b] University of Michigan, Department of Psychology, 530 Church Street, Ann Arbor, MI 48109, USA

Received 15 June 2007; accepted 31 March 2008
Available online 20 June 2008

Abstract

Existing models that integrate emotion and cognition generally do not fully specify why cognition needs emotion and conversely why emotion needs cognition. In this paper, we present a unified computational model that combines an abstract cognitive theory of behavior control (PEACTIDM) and a detailed theory of emotion (based on an appraisal theory), integrated in a theory of cognitive architecture (Soar). The theory of cognitive control specifies a set of required computational functions and their abstract inputs and outputs, while the appraisal theory specifies in more detail the nature of these inputs and outputs and an ontology for their representation. We argue that there is a surprising functional symbiosis between these two independently motivated theories that leads to a deeper theoretical integration than has been previously obtained in other computational treatments of cognition and emotion. We use an implemented model in Soar to test the feasibility of the resulting integrated theory, and explore its implications and predictive power in several task domains.
© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Research on the integration of emotion and cognition has existed for many years (Schorr, 2001). This research has made great strides in establishing that emotion and cognition are, in fact, intimately connected, and several computational models have emerged that embody these ideas (Gratch & Marsella, 2004; Hudlicka, 2004; Neal Reilly, 1996; Ortony, Clore, & Collins, 1988). However, the integrations achieved to date are to some extent incomplete. On the one hand, the claim that cognition is a necessary antecedent to at emotion is well established, and specific cognitive mechanisms that support emotion have even been established (Smith & Kirby, 2001). However, the computational realizations of this integration have largely been pragmatic. Thus, if an emotion theory claims that some cognitive step must take place, such as determining whether a stimulus is relevant to the current goal, then a subsystem is implemented that makes it take place, with little consideration of it overall role in cognition and why it must take place. That is, the link between core cognitive functions and emotion has yet to be fully explored.

Our approach is to start with a theory of cognitive control called PEACTIDM (Newell, 1990; pronounced PEE-ACK-TEH-DIM) and show how a set of emotion theories called appraisal theories naturally fills in missing pieces in PEACTIDM, while PEACTIDM provides the computational structures needed to support appraisal theories. PEACTIDM is a set of abstract functional operations that all agents must perform in order to generate behavior (the acronym denotes these operators, described in detail below: Perceive, Encode, Attend, Comprehend, Tasking, Intend, Decode, and Motor). While PEACTIDM describes the abstract operations, it does not specify the source and types of data that these operations manipulate. We claim that appraisal theories (Roseman & Smith, 2001) provide exactly the required data. Conversely, PEACTIDM provides the functional operations missing from appraisal theories. An important consequence of this integration is that appraisals can be generated incrementally, leading to a time course of emotions. This integration is performed

* Corresponding author.
E-mail address: rmarinie@eecs.umich.edu (R.P. Marinier III).

1389-0417/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2008.03.004

within the Soar cognitive architecture (Laird, 2008), but could equally apply to similar cognitive architectures such as ACT-R (Anderson, 2007). We furthermore show that the integration provides a natural basis for understanding the role of mood and feelings.

The main purpose of this paper is to explore the feasibility and potential value of this integration. Since there are no existing integrations of this kind, a direct comparison to alternative approaches is impossible. Instead, our evaluation focuses on whether the integrated model produces behavior that is qualitatively consistent with PEACTIDM and appraisal theory. We will also address Picard's (1997) list of properties that an emotional system should have (Section 5).

The remainder of this paper is organized as follows: in Section 2, we provide background on cognitive and emotion theories, with a focus on PEACTIDM, Soar and Scherer's (2001) appraisal theory. In Section 3, we describe the unification of these in the context of a model of a simple, short task. In Section 4, we describe a slightly more complex model of an extended synthetic task, and in Section 5, we present an evaluation of that model. Section 6 describes related work, Section 7 describes future work, and Section 8 concludes.

2. Background

In this section, we describe PEACTIDM, a theory of cognitive control, and present background on cognitive theories, particularly Soar, in terms of PEACTIDM. We then present background on emotion theories, and make the connection between PEACTIDM and appraisal theories as complementary pieces of the cognition/emotion integration puzzle.

2.1. Cognitive systems

2.1.1. PEACTIDM: an abstract computational theory of cognitive control

PEACTIDM is a theory of cognitive control where cognition is decomposed into a set of *abstract functional operations* (Newell, 1990). PEACTIDM stands for the set of eight abstract functional operations hypothesized as the building blocks of immediate behavior: Perceive, Encode, Attend, Comprehend, Tasking, Intend, Decode, and Motor. These functions are *abstract* because although many of them may often be primitive cognitive acts, they can require additional processing, whose details are not specified by Newell's theory. PEACTIDM, as Newell described it, was restricted to immediate behavior – tasks with short timescales where interaction with the environment dominates behavior.

We will describe PEACTIDM via illustration with a simple immediate choice response task adapted from a task described by Newell. (As we demonstrate shortly, even a simple example like this can have an emotional component.) In the task, a subject is faced with two lights and two buttons. The lights are both within the subject's fovea. The subject's task is to focus on a neutral point between the lights and wait for a light to come on. When a light comes on, the subject must press the button corresponding to that light. The subject gets feedback that the correct button was pressed by the light turning off in response to the press. The subject's reaction time is the time it takes to turn off the light.

In PEACTIDM, *Perceive* is the reception of raw sensory inputs. In this case, the subject perceives one of the lights turning on. *Encode* is the transformation of that raw sensory information into features that can be processed by the rest of cognition. In this example, a representation is created that indicates one light has come on. *Attend* is the act of attending to a stimulus element. In this case, it is not an overt eye movement but is some type of covert attention that must select the lit light (even though the light is already foveated). *Comprehend* is the act of transforming a stimulus into a task-specific representation (if necessary) and assimilating it into the agent's current understanding of the situation, such as classification or identification. In our example, the subject verifies that one of the two lights has come on (that is, his attention was not drawn by some other stimulus). *Tasking* is the act of setting the task (i.e., the goal) in the internal cognitive state. In our example, Tasking takes place in an earlier cycle before the task begins – the subject is already poised, looking at the lights with a finger ready to press a button and knows which button to press for which light. It is via Tasking that Comprehend knows what to expect and Intend knows what operation to choose based on the input. Given the task and the comprehension of the stimulus, *Intend* initiates a response, in this case, pressing a button. *Decode* translates the response from Intend into a series of motor actions. *Motor* executes the action; in our example, the pressing of the button.

Newell argued that the ordering of PEACTIDM functions is determined largely by the data dependencies between the functions (see Fig. 1). Perceive must occur before Encode, which must occur before Comprehend, which must occur before Intend, which must occur before Decode, which must occur before Motor. In some simple cases, the presence of a stimulus is all that is required for the task, and thus the Encoding step may be skipped. Tasking is the most flexible. In the implementation presented here, Tasking competes with Attend. That is, the agent can either Attend (and thus complete the cycle as shown in Fig. 1), or it can Task (in which case it immediately precedes to Perceive to restart the cycle). An alternative approach has it compete with Intend (see Marinier, 2008).

2.1.2. Approaches to cognitive modeling

Although PEACTIDM describes a set of abstract operations, it does not describe which mechanisms realize these operations and different approaches to cognitive modeling suggest different mechanisms. The *cognitive architecture* approach we pursue here decomposes cognition into more

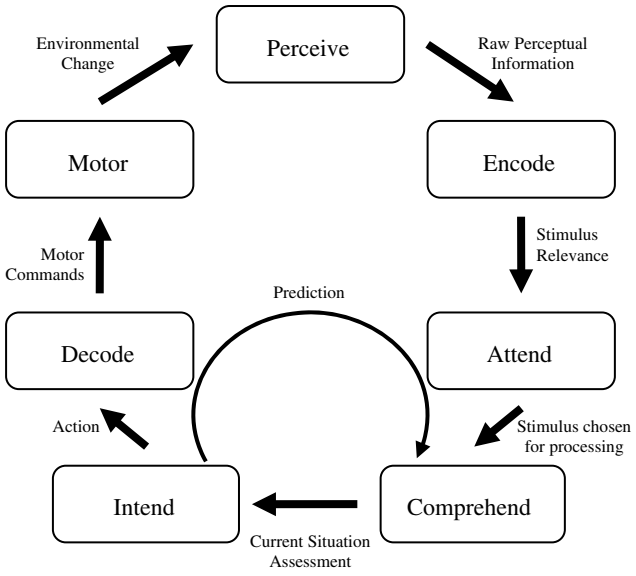

Fig. 1. Basic PEACTIDM cycle. The agent repeats this cycle forever. The output from a step primarily feeds into the next step, but the output of Intend also feeds into the next cycle's Comprehend. Tasking (not shown) competes with Attend. Tasking modifies the current goal, which also serves as an input to the Encode and Comprehend cycles.

primitive computational components that are the building blocks for functional capabilities. The interactions among these components give rise to temporal dynamics within the system. A typical cognitive architecture consists of memories (both long-term and short-term) with different performance characteristics. For example, memories can differ what type of knowledge is stored/learned, how knowledge is represented in the memory, how it is learned, and how it is retrieved. There can also be processing components that combine knowledge, such as to select between alternative interpretations or intentions. Most cognitive architectures also have perceptual and motor systems. Thus, a cognitive architecture provides task-independent structure and subsystems that is shared across all tasks, while using task-dependent knowledge to specialize behavior for a given task. Cognitive architectures are essentially computational systems for acquiring, encoding and using knowledge.

A cognitive architecture implements PEACTIDM by implementing the abstract operations via a combination of its subsystems and knowledge that directs the interactions of those subsystems. We have chosen Soar to realize PEACTIDM, although it should be possible to implement it in other architectures such as ACT-R (Anderson, 2007), EPIC (Kieras & Meyer, 1997), or Clarion (Sun, 2006) (see Marinier (2008) for a description of how PEACTIDM might be implemented in ACT-R).

2.1.3. Soar

Soar is a cognitive architecture that has been used both for cognitive modeling and for developing real-world application of knowledge-rich intelligent systems. Fig. 2 is an abstract block diagram of Soar, which shows the major

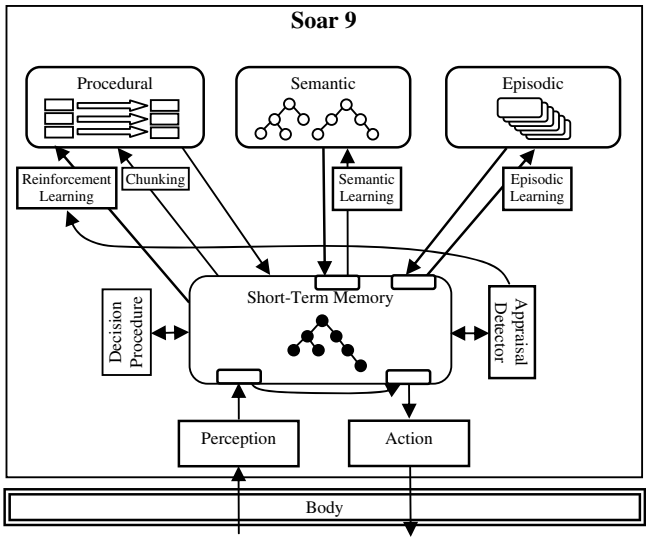

Fig. 2. The structure of Soar.

memories (rounded edges) and processing modules (square edges). In the bottom middle is Soar's short-term memory (often called its working memory). The short-term memory holds the agent's assessment of the current situation, derived from perception (lower middle) and via retrieval of knowledge from its long-term memories. It has three long-term memories: procedural (production rules), semantic, and episodic, as well as associative learning mechanisms. In this work, the semantic and episodic memories are not used, but we will return to them in our discussion of future work. The appraisal detector will be discussed in Section 3.4.

Soar avoids the use of syntax-based conflict resolution mechanisms of traditional rule-based systems by firing all matched rules in parallel and focusing deliberation on the selection and application of *operators*. Proposed operators are explicitly represented in working memory, and deliberation is possible through rules that evaluate and compare the proposed operators. Soar follows a decision cycle (Fig. 3) which begins with an Input phase in which the agent gets input from the environment. This is followed by the Propose phase in which rules fire to elaborate knowledge onto the state, and propose and compare operators. Next, based on the structures created by those rules, Soar selects an operator in the Decide phase and creates a structure in short-term memory representing the chosen operator. This choice may be determined by the comparison knowledge, or it may be random. Once an operator has been selected, rules with knowledge about how to apply that operator can fire. Some of these rules may generate output commands. Finally, Output is processed (e.g., the world is updated in response to an action).

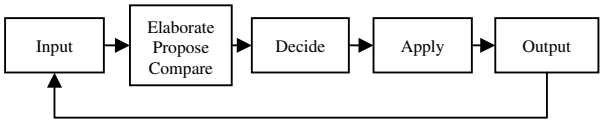

Fig. 3. The Soar decision cycle.

2.1.4. Implementing PEACTIDM in Soar

In this section, we walk through the simple immediate choice response task presented earlier (Section 2.1.1) and describe how it is possible to map PEACTIDM onto Soar. This implementation closely following Newell's (1990) description.

Recall the task situation: the agent is faced with two lights and two buttons; the task is to press the button corresponding to the light that comes on. Before the task even begins, the agent does *Tasking*, which creates a structure in short-term memory describing the goal, which includes a prediction that a light is going to come on. *Perceive* is the reception of raw sensory inputs; in Soar this means that a structure describing which light comes on is created in short-term memory. This structure causes *Encoding* rules in procedural memory to match and generate domain-independent augmentations are added (e.g., the light coming on means the agent can make progress in the task). Rules in Soar fire in parallel, so if there were multiple stimuli, an encoded structure would be generated for each at the same time. *Attend* is implemented as an operator; this is natural since PEACTIDM only allows for one stimulus to be Attended to at a time, and Soar only allows one operator to be selected at a time. Thus, there will be one proposed Attend operator for each stimulus; which one is selected is influenced by the Encoded information. In this task, only one Attend operator is proposed (since there is only one stimulus). *Comprehend* is implemented as a set of operators; exactly how many are required depends on the complexity of the task and situation. In this task, there is only one Comprehend operator, which verifies that the stimulus is what was expected (as determined by Tasking earlier). *Intend* is implemented as set of operators that work together to select a response (in this task, to push the button) and create a prediction of the outcome of that action (in this task, that the light will turn off). In Soar, *Decode* is merely sending the selected action to the output system, and *Motor* is handled by the simulation of the environment.

2.1.5. What PEACTIDM and cognitive architectures provide

PEACTIDM provides constraints on the structure of processing that are more abstract than cognitive architectures like Soar or ACT-R. While Soar and ACT-R specify processing units, storage systems, data representations, and the timing of various mechanisms, they are only building blocks and by themselves do not specify how behavior is organized to produce immediate behavior. PEACTIDM specifies the abstract functions and control that these components must perform in order to produce intelligent immediate behavior.

Some of the key constraints that arise from the combination of PEACTIDM and cognitive architectures are:

- The set of computational primitives that behavior must arise from (Cognitive architecture).
- The temporal dynamics of cognitive processing and behavior (Cognitive architecture & PEACTIDM).
- The existence of core knowledge and structures that must be reused on all tasks (Cognitive architecture & PEACTIDM).

The principle theoretical gain in positing and appealing to a level of analysis at the abstract functional operator level is that it identifies common computational functions across a wide range of tasks. It thus provides a level of description at which a range of regularities may be expressed concerning the nature of these functions. We now exploit this level of description by showing how the inputs and outputs that these operators require implies that they must in fact constitute an affective system of a kind assumed in appraisal theories of emotion.

2.2. Emotion modeling

2.2.1. What can emotion provide?

PEACTIDM and cognitive architectures describe processes and constraints on representation and the timescale of those processes, but they do not describe the specific knowledge structures that are actually used to produce behavior – it is up to the modeler to describe those, and the space of possibilities is large. Consider PEACTIDM: What structures does Encode generate? Given multiple stimuli, what information does Attend use to choose which to focus on? What information does Comprehend generate? What information does Intend use to generate a response? We propose that much of the information required by PEACTIDM is generated by the same processes that generate emotion, and that these processes are, in fact, the PEACTIDM operations themselves. The abstract functions of PEACTIDM need information about relevance, goals, expectations, and so on, and compute them to carry out their functions. The results of these computations, then, cause an emotional response.

2.2.2. Introduction to appraisal theories

The hypothesis that there is a relationship between the way someone interprets a situation (along certain dimensions, such as Discrepancy, Outcome Probability, and Causal Agency) and the resulting emotional response is a defining characteristic of *appraisal theories*. Appraisal theories argue that emotions result from the evaluation of the relationship between goals and situations along specific dimensions (see Roseman & Smith, 2001 for an overview). Appraisal theories are also discussed in Parkinson (2009), Marsella and Gratch (2009), and Reisenzein (2009). For purpose of understanding the functional role of emotion in cognitive architectures, appraisal theories are appealing because they are naturally described at the cognitive level, as opposed to the neurological or sociological levels. Smith and Lazarus (1990) argued that, in general, emotions allow for a decoupling between stimulus and response, which is required to allow organisms to adapt to a broader range of situations. This decoupling, then, meant that more complex cognition was required to fill in the gap. In other

words, complex cognition goes hand-in-hand with complex emotion. Thus, it has been claimed that one of the primary functions of more complex cognition is to support appraisal generation (Smith & Lazarus, 1990).

Appraisal theories fit naturally into our immediate choice response task. When the subject presses the button, he Encodes the state of the light and Attends to it. In the Comprehend stage, he verifies that the light's state matches his prediction. Suppose that after the first several trials, the experimenter disables the buttons so that the light stays turned on even when the correct button is pressed. When the subject Intends pressing the button, he still creates the same prediction – that the light will turn off. When the subject presses the button, though, the light does not turn off. Thus, when the subject gets to the Comprehend step, he will detect a mismatch between the actual state and the expected state.

This mismatch is called Discrepancy from Expectation, and the subject generates a structure to represent it. If the subject has high confidence in an unmet prediction, it might react differently from when the subject has low confidence in an unmet prediction. Thus, when the subject generates the prediction, an Outcome Probably is also generated. In this case, since the subject had no reason to suspect that the light would not turn off when the correct button was pushed, the Outcome Probability was very high.

Since the Discrepancy from Expectation in this case conflicts with the Outcome Probability, we expect the subject would experience surprise. The subject may not even believe what just occurred, and try to press the button again, going through the same steps. However, the second time through, the Outcome Probability is probably lower, and certainly after a few tries, the subject will realize that the button is not functioning. Emotionally, the subject's reaction may vary based on many factors, such as who he thinks is at fault (which we call the Causal Agent). If he thinks he broke the button, he might feel shame. If he thinks he is being thwarted by the researcher, he might feel anger (especially if there was supposed to be some reward based on his performance).

Appraisal theories are complementary to the general cognitive model we described in that they provide a description of the data being processed by cognition. Integration with cognitive architecture can provide the mechanisms and processes that lead to appraisals and which utilize the results of appraisal (e.g., emotions, moods, and feelings; see Sections 2.2.3 and 4.2).

2.2.3. Scherer's appraisal theory

Just as we have chosen to implement our model in a specific cognitive architecture, Soar, we have also chosen a specific appraisal theory to work with: that proposed by Scherer (2001). We do not have a strong theoretical commitment to Scherer model, and we have chosen it largely because of the extensiveness of the theory. Most appraisal theories have six to eight appraisal dimensions, while Scherer's theory has sixteen appraisal dimensions. Thus, in the long run, if we can model Scherer's theory, there is less chance of us missing some important dimension than if we started with a simpler, possibly less complete theory.

Scherer's 16 appraisal dimensions are shown in Table 1. These dimensions are divided into four groups: relevance, implication, coping potential and normative significance. The columns are modal emotions – typical labels assigned to regions of appraisal space close to the sets of values shown.

Scherer's model differs from many appraisal theories in that it assumes a continuous space of emotion as opposed to categorical emotions. Like all appraisal theories, Scherer provides a mapping from appraisal values to emotion labels, but he describes these labels as *modal* emotions – that is, common parts of the emotion space. Given that the majority of existing computational models are categorical (Gratch & Marsella, 2004; Hudlicka, 2004; Neal Reilly, 1996), exploring a continuous model may help clarify the benefits and challenges of such a model. Furthermore, while our theory is continuous, it would be trivial to add categorical labels to regions if desired. Indeed, we introduce a labeling function later that does this (although we use it purely for analysis; see Sections 3.4 and 5.1.1).

Another way in which Scherer's theory differs from most is that he proposes that appraisals are not generated simultaneously. Rather, he claims that appraisals are generated in the order of the groupings given above for efficiency reasons. For example, there is no sense in wasting resources on computing the implications of a stimulus if the stimulus is irrelevant. We will return to this point after we have described our specific model.

Scherer also proposes a process model describing how, at an abstract level, the appraisals are generated and how they influence other cognitive and physiological systems, but it does not provide details of all the data needed to compute the appraisals, nor the details of those computations. Our computational model describes the details. Since the computational details include new constraints on how the model as a whole works, our model differs in some ways from Scherer's theory. This arises in part because of the need to develop a computational model of generation, and also because of the more limited scope of our model. Scherer's theory pays some attention to the physiological and neurological aspects of emotion, but like most appraisal theories, does not include detailed mappings from the theory to specific behavioral data or brain structures. Our model does not include a physiological or neurological model, and does not yet attempt to mode indirect influences on cognition or action tendencies. While these are excellent candidates for future work, our primary focus here is on the generation of appraisals in the context of PEACTIDM, and how appraisals influence behavior; thus, a symbolic cognitive approach is most appropriate.

3. Theory and implementation of integration

The main theoretical proposal is that cognitive and behavioral control, as characterized by PEACTIDM,

Table 1
A mapping from appraisal dimensions to modal emotions with dimensions grouped by function (adapted from Scherer, 2001)

	Enjoyment/ happiness	Elation/ joy	Displeasure/ disgust	Contempt/ scorn	Sadness/ dejection	Despair	Anxiety/ worry
Relevance							
Novelty							
Suddenness	Low	High/med			Low	High	Low
Unfamiliar			High		High	Very high	
Unpredict	Medium	High	High			High	
Intrinsic Pleasantness	High		Very low				
Goal relevance	Medium	High	Low	Low	High	High	Medium
Implication							
Cause: Agent				Other		Other/ nature	Other/nature
Cause: Motive	Intent	Chance/ intent		Intent	chance/neg	chance/neg	
Outcome probability	Very high	Very high	Very high	High	Very high	Very high	Medium
Discrepancy from expectation	Low					High	
Conducive	High	Very high			Low	Low	Low
Urgency	Very low	Low	Medium	Low	Low	High	Medium
Coping potential							
Control				High	Very low	Very low	
Power				Low	Very low	Very low	Low
Adjustment	High	Medium		High	Medium	Very low	Medium
Normative significance							
Internal standards compatibility				Very low			
External standards compatibility				Very low			
	Fear	Irritation/ cold ang	Rage/ hot anger	Boredom/ indiff	Shame	Guilt	Pride
Relevance							
Novelty							
Suddenness	High	Low	High	Very low	Low		
Unfamiliar	High		High	Low			
Unpredict	High	Medium	High	Very low			
Intrinsic Pleasantness	Low						
Goal relevance	High	Medium	High	Low	High	High	High
Implication							
Cause: Agent	Other/natural		Other		Self	Self	Self
Cause: Motive		Intent/neg	Intent		Intent/neg	Intent	Intent
Outcome probability	High	Very high	Very high	Very high	Very high	Very high	Very high
Discrepancy from expectation	High		High	Low			
Conducive	Low	Low	Low			High	High
Urgency	Very high	Medium	High	Low	High	Medium	Low
Coping potential							
Control		High	High	Medium			
Power	Very low	Medium	High	Medium			
Adjustment	Low	High	High	High	Medium	Medium	High
Normative significance							
Internal standards compatibility					Very low	Very low	Very high
External standards compatibility		Low	Low			Very low	High

Those dimensions in italics are not implemented in our current model. Open cells mean all values allowed. Abbreviations: Unfamiliar = unfamiliarity, unpredict = unpredictable, conducive = conduciveness, med=medium, intent = intentional, neg = negligence, ang = anger, indiff = indifference.

requires appraisal information, and that this appraisal information is computed directly by the PEACTIDM operations themselves. The generation of appraisals, and their accompanying emotional responses, then, is a byproduct of the system's normal operation. In this section, we provide the details of the integration of PEACTIDM and

appraisal theory, building on Scherer's (2001) theory as described above (Table 1), though it should be possible to apply other comprehensive appraisal theories in a similar way.

In this section and Section 4, we describe aspects of our theory using examples. In this section, we continue to use the simple choice response task described earlier to give a detailed account of how this integration is realized. Thus, we address how appraisals and emotion are generated and over what time course, how they are represented, how emotion intensity is calculated, and the influence of expectations. Section 4 demonstrates how the model works in a more complex, extended task that we will use to demonstrate additional appraisals and introduce mood, feeling and their behavioral influences.

The simple choice response task follows the steps outlined in Table 2. This version has been slightly extended past our previous description to show what happens immediately following the button push.

To summarize this extended version of the task, the light comes on, and the agent Perceives, Encodes and Attends to the light, and Comprehend verifies that this is what is expected. It then Intends to push the corresponding button. Intend is implemented as a set of operators in Soar that work together to both generate the push button command and create a prediction (that the light will go off). After this command is decoded and physically executed, the light turns off. This change is Perceived, Encoded and Attended, followed by Comprehension. Finally, the agent marks the task complete.

In the process of performing these PEACTIDM steps for this task, appraisal values are generated, which produce an emotional reaction. In this task, only a subset of the appraisals are relevant, namely Suddenness, Goal Relevance, Conduciveness, Outcome Probability, and Discrepancy from Expectation. Fig. 4 shows the relationship between PEACTIDM and appraisal generation and which appraisal information influences which steps in the PEACTIDM process.

Perceive and Encode generate relevance appraisals, which are used by Attend. Comprehend generates assessment appraisals which are used by Intend. Intend generates

Table 2
PEACTIDM steps to simple the simple choice reaction task

PEACTIDM Processing	Notes
	Light on
Perceive	
Encode + Attend	
Comprehend	Verifies prediction
Intend	(to push button) Create Prediction Generate push button command
Decode + Motor	Light off
Perceive	
Encode + Attend	
Comprehend	Verifies prediction
Tasking	(to mark task complete)

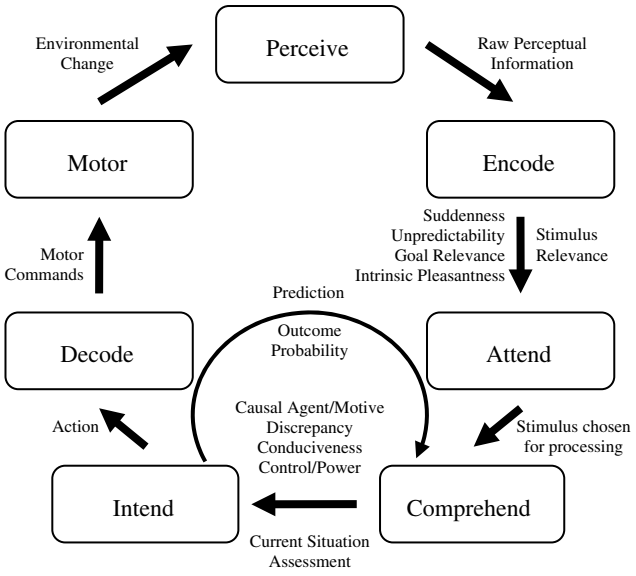

Fig. 4. The PEACTIDM cycle with corresponding appraisals. Suddenness and Unpredictability are actually generated by Perceive, but like other pre-Attend appraisals, are not active until Attend.

the Outcome Probability appraisal, which is used by Comprehend in the next cycle. Tasking (not shown) is influenced by the current emotional state (not shown), which is determined by the appraisals. Critically, our claim is that the PEACTIDM steps require this appraisal information in order to perform their functions, and thus it must be generated by earlier steps.

3.1. Appraisal values

The appraisals differ not only in how they are generated, but also in the types and ranges of values they can have with some appraisal values being numeric, while others are categorical. Table 3 shows the ranges of values we have adopted for the appraisals in our system.

For the numeric dimensions, most existing computational models use the range [0, 1] (e.g., Gratch & Marsella, 2004). The implication is that the 0 end of the range is less intense than the 1 end of the range. For some dimensions, this is true: a stimulus with Suddenness 1 would be considered more sudden that a stimulus with Suddenness 0. For other dimensions, though, being at the "low" end could be just as intense as being at the "high" end. For example, if I pass an exam, I will appraise this as high Conduciveness and have a strong positive feeling. However, if I fail the

Table 3
Appraisal dimensions with ranges

Suddenness [0, 1]	Unpredictability [0, 1]
Goal relevance [0, 1]	Discrepancy from expectation [0, 1]
Intrinsic pleasantness [−1, 1]	Outcome probability [0, 1]
Conduciveness [−1, 1]	Causal agent [self, other, nature]
Control [−1, 1]	Causal motive
Power [−1, 1]	[intentional, negligence, chance]

exam, I will appraise this as very low Conduciveness, (i.e., highly unconducive) and will experience a strong negative feeling. Thus, for these dimensions we use the range [−1, 1] – that is, values near zero (e.g., not very conducive or very unconducive) would have a low impact on feeling, but values near the extremes (e.g., very conducive or very unconducive) would have high impact on feeling.

3.2. Computing the active appraisal frame

In the following sections, we trace the generation of appraisals in our example. To make the calculations easier to follow, we will use extreme values, such as 1.0, for the appraisals, even though less extreme values would be more realistic.

In our example, before the task began (perhaps when waiting for the light to come on), the agent engaged in Tasking which did two things: it created a structure representing the task and a prediction structure that a light will come on. This prediction structure has an associated Outcome Probability appraisal value, which we assume is the extreme value, 1.0. When the light comes on, Perceive generates a value for the Suddenness appraisal, with value 1.0. Then, during Encoding, a structure is created with the following information: which light came on (which is domain-dependent), and whether this stimulus is on the path to completing the task. The fact that a light came on leads to a Goal Relevance appraisal value of 1.0.

The appraisals are stored in an *appraisal frame*, which is the set of appraisals that describe the current situation that the agent is thinking about it (Gratch & Marsella, 2004). Before an agent Attends to a stimulus, there may be several appraisal frames that have been started – one for each stimulus the agent perceives. We call these the pre-attentive appraisal frames. What distinguishes our use of appraisal frames from Gratch and Marsella (2004) is that we use a single active frame to limit which appraisals are generated, whereas they have multiple complete frames; computationally, this makes our approach more efficient.

Attend then uses the available appraisal frames to select the stimulus to attend to. For example, the stimulus that is most Sudden may be preferred. (See the connection between Encode and Attend in Fig. 4.) When a stimulus is Attended, a flag marks the associated appraisal frame as the *active frame*. Once a frame becomes active, several other appraisals can occur. This is in line with our hypothesis that Comprehension follows Attend, and that Comprehension generates the data necessary for further processing (e.g., Intending an action; see the connection between Attend and Comprehend and Tasking in Fig. 4). Specifically, the calculation that the stimulus is on the path to the goal leads to a Conduciveness value of 1.0.

3.3. Sequences and time courses of appraisals

Now that we have described how appraisals are generated, we will discuss the implications of that process on the sequencing and time course of appraisals. Scherer (2001) proposes that the appraisals are generated sequentially because the outcomes of some appraisals obviate the need for others. For example, if none of the relevance appraisals indicates that a stimulus is interesting, then there is no need to continue processing the stimulus. Our model also imposes sequential constraints (see Fig. 4), but for two reasons, one of which is related to Scherer's. Attend will not choose a stimulus unless one of the relevance dimensions indicates that it is interesting, much like Scherer's theory describes. However, additional ordering constraints arise from the flow of data in the model. For example, since Discrepancy from Expectation arises from the Comprehension function, it occurs after the Conduciveness appraisal (which is activated upon Attending). Similarly, the Outcome Probability appraisal is generated in the Intend step, which comes after Comprehension. Thus, while Scherer's argument for sequential appraisal generation centers on efficiency and the wastefulness of generating irrelevant appraisals, our data-driven model extends that to also impose an ordering based on data-driven constraints: the appraisals cannot be generated earlier (regardless of the efficiency). The idea of appraisals being data-driven has been mentioned elsewhere (see Roseman & Smith, 2001, pp. 12–13 for a brief overview of this point), but the idea has been used to argue that appraisal ordering is not fixed at all. Data-driven processing combined with PEACTIDM implies at least a partial ordering.

A corollary to this is that some appraisals take longer to generate than others. In the implementation, all appraisals are generated by rules that test features of the agent's internal state, and thus fire as soon as possible. However, the amount of time it takes to generate the required features varies. As just stated, the Discrepancy from Expectation appraisal rule cannot fire until the required information has been generated by Comprehend (which in turn requires that the Attend operator has been executed). In general, a more complex model might require an arbitrary amount of processing to generate the information necessary so that a Causal Agent appraisal rule can fire, which is consistent with the inference vs. appraisal distinction made by Marsella and Gratch (2009). Thus, the model not only implies partially ordered sequences of appraisals, but it also implies varying time courses for the generation of those appraisals.

3.4. Determining the current emotion

Appraisal theories claim that appraisals cause emotion (see Table 1). Given the theory we have described so far, it may seem that appraisal alone is sufficient. However, as we will see in Section 4, emotion has functional value beyond appraisal, in that it represents situation knowledge in a task-independent form that can be used to influence control and hence behavior. Here we will simply describe the emotion mechanism.

A mechanism called the Appraisal Detector (Smith & Kirby, 2001) processes the active frame to determine the

current emotion. It is via this mechanism that the active frame affects the rest of the system. Emotion theories disagree as to how many emotions a human can have at once. Our current model supports one active appraisal frame at a time, and thus only one emotion (not to be confused with mood or feeling, which are separate; these will be discussed in Section 4). The pre-attentive appraisals generated for the other stimuli do not influence the current emotion in our model.

In many systems (Ortony et al., 1988), the emotion is reported as a label (such as anger, sadness, joy,...) with an intensity. These *categorical* theories of emotion assume that there are a small, fixed number of possible feelings that vary only in intensity. In our model, like in Scherer's (2001) theory that inspires it, each unique appraisal frame corresponds to a unique experience. Categorical, linguistic labels can be generated by segmenting the space of appraisal frames, and we do this for our own analytical purposes. However, the current model does not use these labels, and even if it did, at best such labels would be a model of how an individual in a particular culture might label the emotions. For example, in the current problem, since Conduciveness and Goal Relevance are positive, and other appraisals such as Causal Agent are not being considered (which would lead to Pride), the agent's current emotion would correspond to joy. The actual representation is the active appraisal frame: Suddenness = 1.0, Goal Relevance = 1.0, Outcome Probability = 1.0, and Conduciveness = 1.0.

3.5. Calculating intensity

In addition to determining an appraisal as a point in a multi-dimensional space (or as a category), the system must also determine the *intensity*. Intensity is important because it summarizes the importance of the emotion, and thus indicates to what degree it should influence behavior. Emotions with low intensity are likely to be caused by less important stimuli than emotions with high intensity. In this section, we briefly present the intensity function; see Marinier and Laird (2007) for details on the derivation of this function.

Overall our approach combines the numeric dimensions of the active appraisal frame to form a single numeric intensity value; since the categorical dimensions are non-numeric, they do not participate in the intensity calculation. There are many ways to produce an intensity value from a frame, and although there is little theory or empirical evidence to guide us, we define three general criteria for an intensity function:

(1) Limited range: intensity should map onto [0, 1]. This is common to most existing theories.
(2) No dominant appraisal: no single appraisal value should dominate the intensity function; each should contribute to the result but no single value should determine the result. This criterion eliminates a commonly used basis for combination: multiplication (e.g., Gratch & Marsella, 2004). One critical problem with multiplication is that if any dimension has a zero value, then the intensity will be zero, regardless of the other values.
(3) Realization principle: expected stimuli should be less intense than unexpected stimuli (Neal Reilly, 2006). This is in contrast to Gratch and Marsella (2004) where intensity is maximized when the likelihood is 1.

Our intensity function has two parts: a *surprise factor* that takes into account how expected or unexpected a stimulus is based on the Outcome Probability and Discrepancy from Expectation dimensions, and an averaging part that incorporates the rest of the numeric appraisal values. The intensity equation is

$$I = [(1 - \mathrm{OP})(1 - \mathrm{DE}) + (\mathrm{OP} \cdot \mathrm{DE})] \cdot \frac{S + \mathrm{UP} + \frac{|\mathrm{IP}|}{2} + \mathrm{GR} + \frac{|\mathrm{Cond}|}{2} + \frac{|\mathrm{Ctrl}|}{2} + \frac{|P|}{2}}{\mathrm{num\_dims}}$$

where OP is the Outcome Probability, DE is the Discrepancy from Expectation, S is the Suddenness, UP is the Unpredictability, IP is the Intrinsic Pleasantness, GR is the Goal Relevance, Cond is the Conduciveness, Ctrl is the Control, P is the Power, and num_dims is the number of dimensions included in the average (7, if all dimensions have values). In those cases where one or more values for appraisals in the averaging part of the equation are missing (as in our current simple choice reaction task example), the average is taken over the values that are present. If either Outcome Probability or Discrepancy from Expectation is missing, then the present value is multiplied by the averaging part (in this model, the Outcome Probability is always present in an active appraisal frame since there is always a prediction).

The intensity function is biased so that some classes of emotions are inherently more (or less) intense than others. For example, the emotions that Scherer's theory would label as Boredom/Indifference are composed of low values for most dimensions combined with high outcome probability and low discrepancy, resulting in low intensity (see Table 1 for Scherer's mapping from appraisals to emotions). On the other hand, Scherer's Rage/Hot Anger emotions are composed of mostly high values, with high outcome probability and high discrepancy, resulting in high intensity. This is congruent with many circumplex models of emotion (Yik, Russell, & Feldman Barrett, 1999), which also propose different intensities for different emotions, suggesting a bridge between circumplex models and appraisal models.

3.6. Modeling the task

Returning to our example, the intensity of the joy following the light coming on is Outcome Probability multiplied by the average of Suddenness, Goal Relevance, and Conduciveness. Since these all have value 1, the intensity is 1. Fig. 5

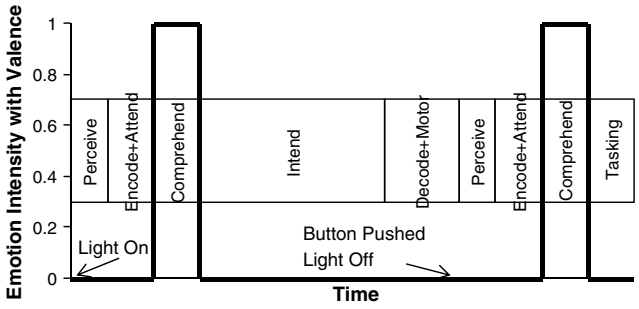

Fig. 5. The task as split into PEACTIDM stages with the signed emotion intensity at each point in time.

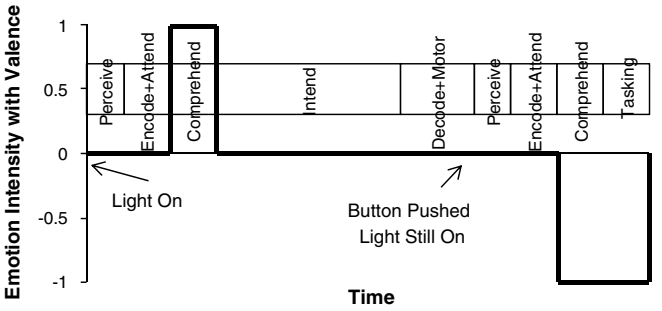

Fig. 6. The revised task as split into PEACTIDM stages with the signed emotion intensity at each point in time.

shows the entire task in terms of the PEACTIDM stages with the emotion intensity at each point in time.

Next, the agent verifies the prediction in the Comprehend step. Recall that the prediction was created before the task began, and it said that a light would come on. The prediction was accurate, so a value of 0 is generated for Discrepancy from Expectation. This causes the intensity of the emotion to drop to 0 because the surprise factor of the intensity is 0 (we might now call the emotion boredom).

Following Comprehend, the agent Intends to push the button. As described earlier, this causes the architecture to generate a prediction that the light will go off when the button is pressed, and it generates the command to push the button. The prediction replaces the previous prediction (that the light would come on) and has a new Outcome Probability associated with it (again, let's assume it is 1). This is followed by Decode and Motor with the result that the button is pushed and the light turns off. This change is Perceived, Encoded and Attended with appraisals generated as before, again resulting in a positive emotion with an intensity of 1. Comprehend confirms the prediction, causing the intensity to return to 0. Finally, Tasking marks the task structure as complete.

3.7. The revised task

When the world behaves as expected, there is very little to get excited about. Emotional reactions are often strongest when unexpected things occur. To explore this, we revised the task so that the light does not turn off when the button is pushed. How does this change the appraisals? The first part of the task (up to the pushing of the button) is exactly the same so that the Suddenness and Goal Relevance appraisals have values of 1, just like before. However, now when the button is pushed, nothing happens, so that when the stimulus (the light) is Attended to, Conduciveness is −1 because the stimulus is not on the path to the goal, as shown in Fig. 6. The intensity of the emotion is still 1, but the valence is negative (because Conduciveness is negative). Our labeling function (Section 5.1.1) calls this appraisal frame Displeasure. Comprehend determines that the prediction was inaccurate, resulting in a Discrepancy from Expectation value of 1. Thus, whereas before the intensity returned to 0 at this point, it now stays at 1, and thus the negative emotion persists (see Fig. 6). We can only speculate at what would happen next, since the situation is presumably not covered by the task instructions; in our version, the agent still does Tasking and marks the task as complete.

3.8. Discussion of the model

The emotional reaction of an agent to the task depends on at least two factors: to what extent the things occur as the agent has predicted them to, and what is at stake for the agent. In Fig. 5, the agent has very brief reactions to the stimulus (in Soar, on the order of 50 ms), which immediately go away when the agent realizes that the results are consistent with its expectations. This demonstrates how incrementally generated appraisal information leads to the emotion time courses. In Fig. 6, when the outcome is unexpected, the agent's reaction is prolonged. Thus, even for a mundane task like pushing a button, emotional responses are possible. In the example, the appraisal values were extreme for demonstrative purposes, which would reflect a situation in which the consequences of the agent's actions are extremely important – such as the World Championship of button pushing, or if a large amount of money is riding on the agent's performance. One has only to watch TV game shows where the only action is choosing a box to open to see examples of extreme emotional responses for mundane actions. To emulate mundane button pushing, lower appraisal values would be used, which would result in little emotional reaction.

3.9. Summary

In this section, we demonstrated the integration of PEACTIDM and appraisals in our implementation. This included many details that go beyond PEACTIDM and appraisal, including value ranges for appraisals, active appraisal frames, and the calculation of intensity. An additional avenue of inquiry is the relationship between the agent's performance on this task and human data, and the impact of Soar's chunking mechanism. Those issues are described in Marinier (2008).

The next section describes the model in the context of a task that involves multiple actions over time. At the end of

that section will be a discussion of some of the implications of the model which apply equally well to this simple model, but which the reader may find easier to appreciate in the more complex context. Hence, that discussion is delayed until then.

4. A model in a more complex, extended task

In the previous section, we described the integration of PEACTIDM and appraisal theory in Soar in the context of a very simple task. In this section, we extend that model to a more complex (but still fairly simple) extended task that utilizes more appraisal dimensions. Unlike the previous task, this task may take an arbitrary number of PEACTIDM "cycles" to complete. This raises new issues, such as how previous emotions affect new emotions, and the role of Tasking when the ongoing task may be viewed as different subtasks. Addressing these issues will allow us to address qualitative questions such as, does the model produce coherent, useful behavior in the long term? Do the appraisals affect behavior and vice versa? Do appraisals have a reasonable (if not human-matching) time course? These and other questions will be addressed in the evaluation (Section 5).

For an ongoing task, we have chosen a simple Pacman-like domain called Eaters (Fig. 7a) that eliminates complexities of real-world perception and motor actions, while supporting tasks that although simple, allow for a range of appraisals and emotions. Eaters is a 2-D grid world in which the agent can move from square to square except where there is a wall. The agent can sense the contents of the cells immediately to its north, south, east and west. The agent's task is to move from its starting location to a specified goal location. This may not always be possible, in which case an intelligent agent should choose to give up so it can move on to other tasks. The task ends when the agent notices it has achieved the goal or when it gives up.

In terms of PEACTIDM, the agent will need to Perceive its surroundings, including information about what lies in each direction (e.g., walls, open spaces), create structures representing the encoded form of the input (e.g., some direction is passable and whether moving in that direction leads closer to the goal), Attend to one of the encoded structures, Comprehend that structure in terms of its current understanding of the situation (e.g., is the situation what the agent predicted), Intend an action if possible (e.g., if the Attended structure can be acted upon to get closer to the goal), and then perform the Intended action (via Decode and Motor). Tasking will play a role when the agent is stuck; for example, it may need to create a subtask to circumvent a wall, or to give up.

In appraisal theory terms, each choice point (e.g., what to Attend to, what to Intend, when to give up) will be guided by emotional information. Thus, the steps preceding these choice points must generate the appraisals that, directly or indirectly, influence the choices to be made.

What follows are the details of how each PEACTIDM function is implemented in this model, including how the appraisals fit in.

4.1. PEACTIDM in the Eaters domain

This section describes how PEACTIDM as implemented in Soar is used to perform the Eaters task. Some aspects of these phases are domain-specific (e.g., the stimuli and actions), but most of the core processing (Encode, Comprehend, Tasking) is general and taken directly from the previous model.

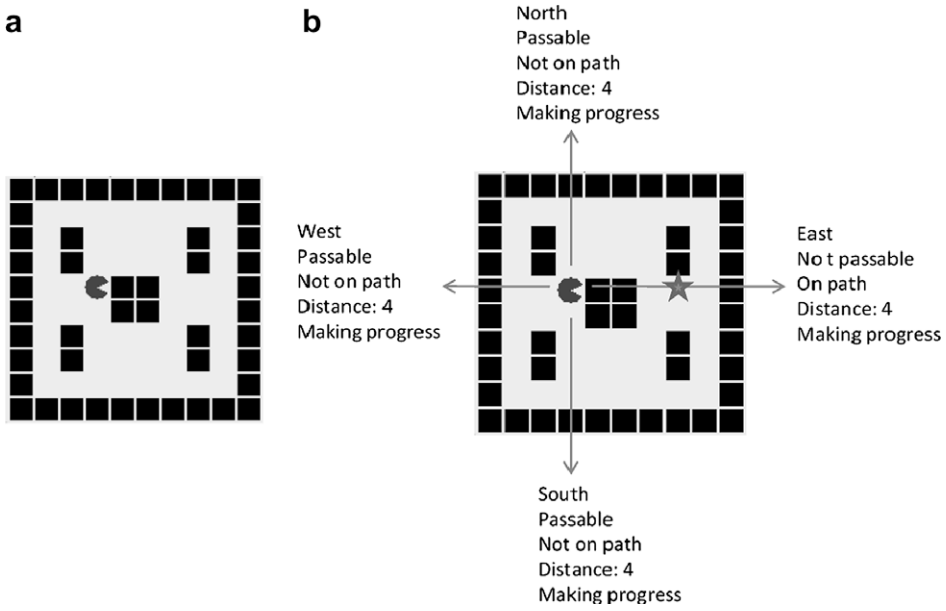

Fig. 7. (a) A screenshot of eaters. The agent is the Pacman-like figure at location (3, 4), walls are black cells, and open spaces are light-colored cells. (b) Encoded structures for each stimulus. The star shows the goal location.

4.1.1. Perception and encoding

Perception and Encoding generate structures that lead to relevance appraisals used by Attend to determine which stimulus to process. We do not directly model the Perceive function. The Eaters environment provides symbolic inputs to the Soar agent. Each direction (north, south, east, and west) is considered a stimulus; thus, a separate structure is Encoded for each direction, which includes information such as whether the direction is passable, whether it is on the path to the goal or not, the distance to the goal, and whether the agent is making progress. The distance to the goal is an estimate based on Manhattan distance and may be incorrect if there are walls between the agent and the goal. If the agent is at a goal location, it will have a separate Encoded structure for the goal completion. The Encoded structure is fairly general – any task in which there is a path to the goal that can be blocked and where there is an estimate of distance to the goal can be Encoded in this way.

Fig. 7b shows an example that will be used throughout the rest of this section. The goal is for the agent to reach location (7, 4) (marked by the star) and the agent has moved from the west. The agent will have four encoded structures, one for each cardinal direction. The north, south, and west structures will be marked as passable, directly off the path (since those directions will increase the distance to the goal), and at a distance of 4 from the goal. The east structure will be marked as impassable but directly on the path to the goal.

Relevance appraisals are generated directly from these Encoded structures. The north, south, and east stimuli have some Suddenness, whereas the west stimulus has no Suddenness (since the agent just came from there). In any environment, the agent will likely have some general expectations about what things to expect, and our agent expects there not to be many walls in the world. Thus, the north, south and west stimuli have low Unpredictability, but the east stimulus has a high Unpredictability. Our agent is also averse to walls (since they only ever get in its way). Thus, it finds them Intrinsically Unpleasant giving the east stimulus a low Intrinsic Unpleasantness value. Finally, since the east direction is on the path to the goal, it is highly Goal Relevant, but the other stimuli are not (Fig. 8a). Note that, in this model, only one goal or subgoal is active at a time, and thus Goal Relevance is computed with respect to that goal.

4.1.2. Attending

In general, the agent wants to make progress towards its goal, so stimuli that are Goal Relevant should given priority. However, Sudden or Unpredictable stimuli may also require attention, since these may be signals of danger or opportunity that needs to be dealt with. This is essentially an exploit versus explore tradeoff. Finally, stimuli that are intrinsically pleasant or unpleasant (independent of the current goal) may also deserve attention. In this model, each stimulus is appraised along the Suddenness, Unpredictability, Intrinsic Pleasantness, and Goal Relevance dimensions, determining the appraisal frame (Fig. 4).

In this model, the selection of which stimulus is Attended to is a weighted random choice, with weights determined by the values of the appraisals just discussed. Since unusual stimuli are more likely to be worthy of Attention, as described above, appraisals with more extreme values lead to larger weights; that is, more interesting stimuli are more likely to be Attended to. Thus, the appraisals provide a task-independent language for knowledge that can influence control.

In our example, the north and south Attend proposals have moderate weights, whereas the west Attend proposal has a slightly lower weight (since its Suddenness is lower). The east Attend proposal has a higher weight because it is

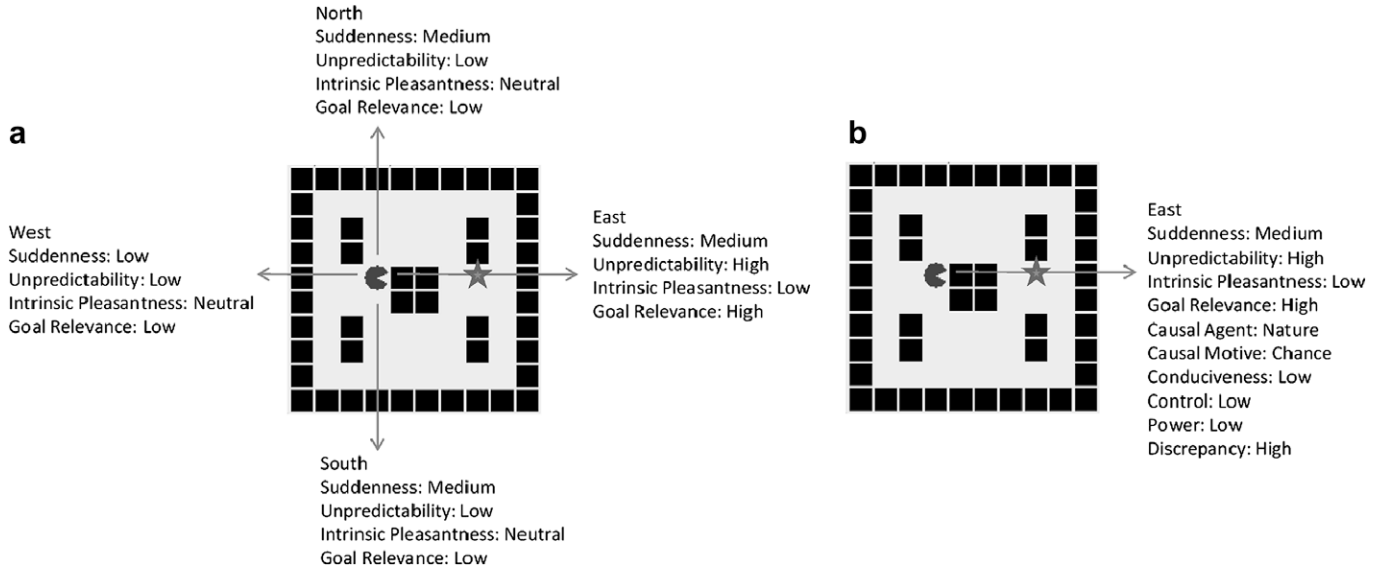

Fig. 8. (a) Pre-attentive appraisal frames for each encoded structure. (b) The agent attends East, making that appraisal frame active. The Comprehend function adds to this frame. The agent decides to Ignore this stimulus.

on the path to the goal, leading to an appraisal of Goal Relevance, and it has a wall, which is Intrinsically Unpleasant. Thus, the agent is most likely to Attend east.

4.1.3. Comprehension

Next, the agent performs the Comprehend function, which adds several additional appraisal values to the active frame (Fig. 8b). The agency of the stimulus is determined (in this model, "nature" is always the Causal Agent and "chance" is always the Causal Motive). The Conduciveness is also determined – if the stimulus direction is passable and on the path to the goal, it has high Conduciveness, whereas if it is off the path or blocked it has low Conduciveness. The Control and Power appraisals are also generated – if a stimulus direction is passable, Control and Power are rated high, whereas if the direction is impassable, Control and Power are low. While this domain is very simple, and thus the generation of these appraisal values is very simple, a more complex domain would potentially require arbitrary processing to determine values for any of these appraisals. We will not consider such extended processing here.

In our example, since the agent is Attending to the east stimulus, which is impassable but on the path to the goal, it will generate appraisals of low Conduciveness, low Power, and low Control (since it cannot walk through walls). Causal Agency and Motive are "nature" and "chance", as noted above.

As in the previous model, the agent then verifies the stimulus via comparison of the current stimulus to the current prediction (as generated by the previous Intend) leading to the generation of the Discrepancy from Expectation appraisal. If the stimulus is a match, then the Discrepancy from Expectation appraisal is low; if there is not a match, then the Discrepancy is high.

Unlike the previous model, once a stimulus has been verified, the agent performs another Comprehend step that determines if further processing is warranted. This gives the agent a chance to "back out" if it determines that processing should not proceed. That is, the agent answers the question, can additional processing of this stimulus lead to an action that helps me? The agent uses a heuristic called *dynamic difference reduction* to make this choice. Difference reduction (Newell, Shaw, & Simon, 1960) attempts to take internal processing steps to reduce the difference between the current state description and the goal state description. Dynamic difference reduction (Agre, 1988) takes the steps in the world to avoid the need for increasing amounts of memory to track one's imaginary progress. Thus, difference reduction leads to plans whereas dynamic difference reduction leads to actions. In our model, if a stimulus can be acted upon (i.e., it is associated with a passable direction) and it does not lead directly away from the goal, then Comprehension is complete and the agent acts upon it (it does the Intend function). Otherwise, the agent chooses a second Comprehend operator, Ignore. Ignore marks the stimulus as processed and allows control to return to Attend, which will choose another stimulus to process from the remaining stimuli as above. This deactivates the appraisal frame for the Ignored stimulus.

In our example, the agent is Attending east, which is a wall. Comprehend will find a mismatch (since our simple model almost always predicts a passable route to the goal). This will trigger an appraisal of high Discrepancy from Expectation, which is added to the current frame. Since there is a wall, the agent cannot directly act upon the stimulus, so it then Ignores it. In fact, the agent is trapped by its goal in this case. As it Attends and Comprehends to each stimulus, it will find that the remaining stimuli lead away from the goal. Thus, Ignore will eliminate all of the remaining stimuli.

4.1.4. Tasking

When the agent has no options left, it is forced to engage in Tasking. This is an addition to the previous model which did not engage in Tasking during the task itself (only before the task began and at the very end). Generally speaking, Tasking is about managing goals (e.g., creating goals, giving up on goals, etc.). In this case, the agent creates a subtask to get around the blockage. In general, there are at least two types of goals. One type is abstract – the goal cannot be acted upon directly and must be broken down into more concrete components (perhaps many times) until it is in a form that can be directly acted upon. For example, the goal "Go to Work" is very abstract, and must be broken down to something that can be directly executed, such as "take a step". The other type is concrete – the goal can be acted upon directly. This is the form of goals in this model. When the agent temporarily retasks itself for the purpose of making progress on its original goal, we call this subtasking, and we call the new goal structure a subtask.

The goal that the agent cannot make progress on is to go to (7, 4). The reason that the agent is stuck on this goal is that its control knowledge and task formulation are too restrictive. Movement in any available direction will take it further from the goal, which violates its dynamic difference heuristic. In order to move around the blockage, it needs to temporarily get further away from the goal. Thus, the agent needs to retask and create a goal that is less constraining, allowing it to get further from the main goal, but without violating its constraints in the new goal. The agent does this by defining the step it would ideally take – in this case, it would ideally move east to $x = 4$. It sets this as its new subtask. That is, there is no constraint in the y (north-south) direction.

When an agent creates a subtask, it records information that gives it some idea of whether it is making progress or not. Specifically, it records the distance to the parent task (goal) at that time. It also tracks the minimum distance it has ever been to the goal upon entering a subtask. If the current distance to the goal is less than the minimum distance to the goal, then the subtask is considered a "good" subtask – that is, the agent knows that, even though it has

to retask, it is making progress towards the goal. If the distance to the goal is not reduced, then the subtask is considered a "bad" subtask – that is, the agent cannot tell if it is actually making progress by retasking. The Encode function adds this good/bad subtask information to each Encoded structure, and this information influences some of the appraisals. In this model, the Conduciveness appraisal is more positive in good subtasks.

As alluded to above, once the agent has this new subtask, the Encoded stimuli are regenerated (since there is a different context for them now) and the agent can then re-Attend to the stimuli to see if any are now suitable. The agent can theoretically create an arbitrary number of nested subtasks this way, but for the current task it only needs one at a time (although it may create several in the course of completing the goal).

In our example, this is the agent's first subtask, so it defaults to a good subtask. The agent might still Attend to the east stimulus first and ignore it again, but when it Attends to, for example, the north stimulus, it will find that it is no longer directly off the path to the subtask. Instead it is now a sideways move (since it neither gets it closer to nor further away from $x = 4$). Thus, the agent determines that this stimulus can be used for Intention processing (Fig. 9a).

4.1.5. Intending

Once the agent has found a stimulus it can act upon, it performs the Intend function, which is also implemented as a Soar operator. As in the previous model, Intend proposes moving in the direction of the stimulus. It also creates a new prediction structure – namely that the next stimulus direction will be passable and on the path to the goal (Fig. 9b) in this model, the agent is always optimistic in this way). If the agent is currently one step away from the goal, then it creates a goal achievement prediction. Along with the prediction, the agent also generates an Outcome Probability appraisal. As before, the Outcome Probability is tied to the prediction, and thus all appraisal frames in the situation that results from an Intend will inherit this same Outcome Probability (Fig. 4).

In our example, Intend proposes moving north. The Intend operator sends a command to the environment to move north, and also creates a prediction. Since it is pursuing a subtask, the agent is less confident of its predictions, so it only rates the Outcome Probability of this prediction as moderate.

4.1.6. Decode and motor

We do not directly model the Decode and Motor functions. The model uses Soar's standard method of communicating an action command to the simulated environment, which then executes it, leading to a new input state. For simplicity, in the model presented here, actions never fail (e.g., if the agent Attends to a wall, it will Ignore it instead of trying to move into it). However, more recent work on learning (Marinier & Laird, 2008) does allow action failures.

4.2. Emotion, mood, and feeling

In the previous model, we described how active appraisal frames become emotions. That is still true in this model. However, since the agent behaves over a long period of time in this task, the question naturally arises, how do emotions affect each other over time? In this section, we will introduce mood and feeling. The functional aspects of these will be discussed in Section 4.3.

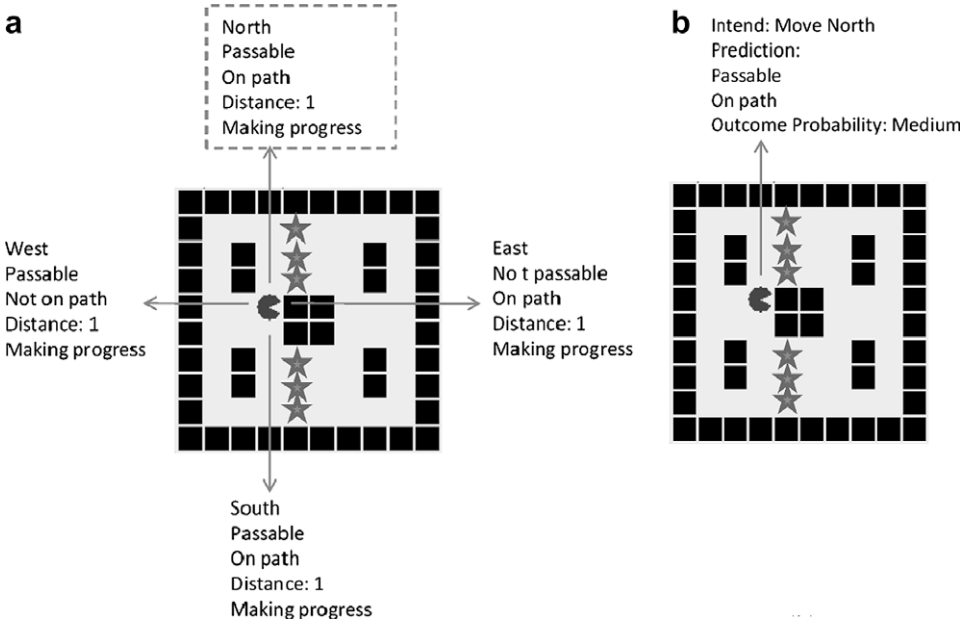

Fig. 9. (a) The agent creates a subtask to get around the blockage. The stars show the possible locations that would solve the subtask. This causes new encoded structures to be created. The agent Attends north. (b) The agent Intends moving north. It creates a prediction of the next stimulus it will see.

Recall that some existing computational models attempt to address the issue of how an emotion affects a succeeding one (see Section 6). Still, these models, and most theories, do not make an explicit distinction between emotion, mood and feelings; some only describe emotion (Hudlicka, 2004), some only describe emotion and mood (Gratch & Marsella, 2004) and some describe emotion, but mood only vaguely (Smith & Lazarus, 1990). One existing distinction made between emotion and mood is in terms of timescale: emotions are short-lived while moods tend to last longer (Rosenberg, 1998). Some physiologically-oriented theories of emotion (Damasio, 1994; Damasio, 2003) distinguish between emotions and feelings: emotions have some impact on physiology, and the agent perceives or *feels* these changes, called the agent's feelings. That is, feelings are our perception of our emotions.

This distinction between emotion, mood and feeling is not universally accepted; indeed, what processes and phenomena are considered "emotional" is a subject of considerable debate. In our model, the specific labels are less important than the computational processes, structures and connections that make up the model as a whole. For example, Frijda, Kuipers, and ter Schure (1989) consider action tendencies to be part of emotion, whereas in our model we have action tendencies separate from emotion. Nevertheless, since the architecture supports the generation of action, and we have added the ability to generate emotion, mood, and feeling, the mechanisms are in place to allow an integration of these with action. Indeed, action is partially influenced by feeling in the present model (see Section 4.3). That these phenomena are inextricably bound is not debated; how we choose label them is an expository convenience.

In our model, we maintain a distinction between emotion and feeling, and also introduce mood. Emotion is the currently-active appraisal frame. In our model, we use a simple model of mood, where mood is a weighted average-like aggregation over past emotions computed at the individual appraisal level, so that mood is represented as an appraisal frame. This initial model of mood captures some of the time course and interactions among emotions, while ignoring many of the complexities of a more complete model of mood. Feeling is the combination of emotion and mood, represented as an appraisal frame, augmented by an intensity. Thus, in the previous model, what we reported as the agent's emotion with intensity (e.g., joy, 1.0) is actually the agent's feeling and feeling intensity. Since feeling is represented using an appraisal frame, the intensity calculation we proposed previously (Section 3.5) still applies. For more details, see Marinier and Laird (2007).

Fig. 10 shows how the agent generates an appraisal frame (its emotion), which interacts with another appraisal frame (its mood) to generate its perceived appraisal frame (its feeling). Given a feeling frame, the system calculates the intensity of that feeling (using the method described in Section 3.5). The mood starts out neutral (i.e., all zero values). To model the influence of emotion on mood, the

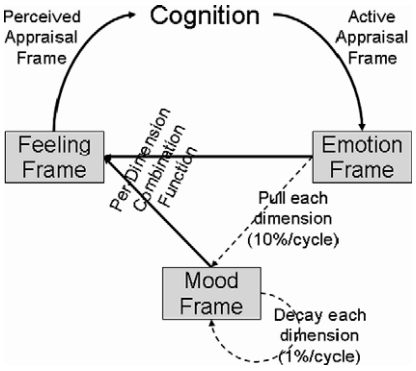

Fig. 10. An emotion frame influences and combines with the mood frame to produce the feeling frame, which is perceived by the agent.

mood "moves" towards the emotion each time step. In the current model, we have adopted a simple approach where the mood moves x% (our current experimental value is 10%) of the distance along each dimension towards the emotion in each cycle. Additionally, the system decays mood by y% (experimental value is 1%) each cycle. Thus, each emotion influences mood for a theoretically infinite amount of time, but the magnitude of the influence decreases exponentially with time. Therefore, if there were no influence of emotion, mood would eventually become neutral.

We make the simplifying assumption that the dimensions are independent, so our combination function takes as input a particular dimension from the mood and emotion frames to produce the corresponding dimension of the feeling frame. This function is applied to each dimension of the frames.

We used the following criteria to create our combination function. Some of these criteria are derived from prior work (see Marinier & Laird, 2007 for details):

(1) Distinguishability of inputs: Large input ranges should have large output ranges. Capping of extreme values may be necessary, but it should have minimal impact.
(2) Limited range: $C(v_{\text{emotion}}, v_{\text{mood}})$ should be between the input with the maximum magnitude and the sum of the inputs.
(3) Non-linear: For small inputs, C is nearly additive, but for large inputs, C is closer to a max. Put another way, for small values the derivative of C can be close to 1, but for large values, the derivative of C should be closer to 0.
(4) Symmetry around 0: $C(x, 0) = C(0, x) = x$. If the mood or emotion input is 0, then the other input dominates. If they are both zero, then the result should be zero.
(5) Symmetry of opposite values: $C(x, -x) = 0$. The mood and emotion can cancel each other out.
(6) Symmetry of all values: $C(x, y) = C(y, x)$. The mood and emotion have equal influence on the feeling.

Using these criteria, we derived the following function (based on Neal Reilly's (2006) function):

$$C(v_{\text{mood}}, v_{\text{emotion}}) = 0.1 \cdot \text{Sign}(S) \cdot \log_b(|S + \text{Sign}(S)|)$$

$$\text{where} \quad S = \sum_{v=v_{\text{mood}}, v_{\text{emotion}}} \left(\text{Sign}(v) \cdot (b^{10 \cdot |v|} - 1)\right)$$

$$\text{and} \quad \text{Sign}(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ -1 & \text{else} \end{cases}$$

$$\text{and} \quad b = \begin{cases} e & \text{if } \text{Sign}(v_{\text{mood}}, v_{\text{emotion}}) = 1 \\ 1.1 & \text{else} \end{cases}$$

$$\text{If } C(v_{\text{mood}}, v_{\text{emotion}}) > 1 \text{ then } C(v_{\text{mood}}, v_{\text{emotion}}) = 1$$

$$\text{If } C(v_{\text{mood}}, v_{\text{emotion}}) < -1 \text{ then } C(v_{\text{mood}}, v_{\text{emotion}}) = -1$$

The combination function, together with the intensity function we presented earlier, can sometimes lead to unexpected results. Even though the combination function has a building effect (i.e., if the inputs have the same sign, the magnitude of the result will be at least as large as the magnitude of the largest input), this will not necessarily result in a higher the intensity for the feeling. Given the way Outcome Probability and Discrepancy from Expectation influence intensity via the surprise factor, even if both of those values go up, the intensity may actually go down.

Unlike other models (Gratch & Marsella, 2004; Hudlicka, 2004; Neal Reilly, 1996) the mood and feeling processes do not combine emotions; they combine individual appraisals. This could lead to unexpected feelings. For example, an emotion best described as elation–joy combined with a mood best described as anxiety–worry can result in a feeling best described as displeasure–disgust. This is an interesting prediction of the model that we have not yet investigated.

4.3. The influence of emotion, mood and feeling upon behavior

Feeling adds knowledge to the state representation in a task-independent format that combines representations of current (emotion) and past (mood) situations, and thus is more general than emotion or mood alone. Feeling can be used to guide control, and thus it can influence behavior. Task-dependent representations can still influence behavior both directly (as in how the agent might choose to cope with its feelings in a particular domain) and indirectly (in that appraisals can be generated from task-dependent representations). Emotion theories describe a number of influences of emotion, mood, and feeling, including effects on cognitive processing (Forgas, 1999) and coping (Gross & John, 2003), and integration with action tendencies (Frijda et al., 1989). Our current approach is very simple, included to demonstrate the possibility of feelings influencing behavior and focusing on one aspect of coping: coping by giving up on goals.

Most AI systems, when faced with a difficult or impossible task, have no way to recognize that they should give up and will work on the problem until all resources are exhausted. By providing emotional feedback, our model allows the agent to detect that it is not making progress towards the goal, and thus it can choose to discard that goal (possibly so it can move on to another goal or stop wasting resources). This behavior could be accomplished without emotions, moods, and feelings, but they provide a natural way to achieve this.

In our model, when the agent fails to make direct progress, it will form a subtask. While pursuing a subtask, the agent can choose to give up if its current feeling of Conduciveness is negative. Giving up is another form of Tasking – it removes the current goal. As this feeling intensity increases, the agent is exponentially more likely to give up. Mood plays a role here by tempering or enhancing the current emotion. Thus, if things are going well (mood is positive) but the agent experiences a momentary setback (emotion is negative), the overall feeling intensity will be lower, making giving up less likely. If things have been going poorly, however, the setback will build on that, resulting in a more intense negative feeling, making giving up more likely. The option to give up is in competition with other activities in the subtask, specifically attending to possible directions in which it can move. That is, the agent still makes a weighted random choice, with giving up being an option whose weight is exponential in the magnitude of the negative feeling intensity. As the agent eliminates more of its Attend options (by Attending to and then Ignoring them), it becomes more likely to give up (since there is less competition from other Attend proposals).

While the current model only has this single direct influence of feelings on behavior, each appraisal of each stimulus has an indirect influence. As described above, at the Attend stage, the pre-attentive appraisals influence where attention is focused next. Furthermore, past appraisals influence the current feeling via mood, and thus indirectly influence the agent's decision to give up or not.

5. Evaluation

What kind of evaluation is appropriate for this model? Clearly, given the computational nature of the system, it is possible to generate quantitative results. However, given the lack of human data or existing systems to compare to, these results can only be used to support claims about the system itself, as opposed to a comparison.

First we consider Picard's (1997) properties that an emotional system should have:

(1) Emotional behavior: system has behavior that appears to arise from emotions.
(2) Fast primary emotions: system has fast "primary" emotional responses to certain inputs.
(3) Cognitively generated emotions: system can generate emotions, by reasoning about situations, especially as they concern its goals, standards, preferences, and expectations.

(4) Emotional experience: system can have an emotional experience, specifically cognitive and physiological awareness and subjective feelings.
(5) Body–mind interactions: system's emotions interact with other processes such as memory, perception, decision making, learning, and physiology.

We begin with 3 (cognitively generated emotions). The system has this property as it uses cognitively generated appraisals as the basis for its emotions. Similarly, the system exhibits 2 (fast primary emotions) because the system generates appraisals beginning at the Perception and Encoding phases, and those become active at the Attend phase. While some have argued that appraisals are "too cognitive," and thus cannot be used to generate fast emotional responses (Zajonc, 1984), Soar naturally supports this fast appraisal generation, so long as no significant inference is required (Marsella & Gratch, 2009). Indeed, one implication of the Scherer (2001) theory is that the relevance appraisals (suddenness, unfamiliarity, unpredictability, intrinsic pleasantness, goal relevance) are generated very early, and our system reflects that. Moreover, as soon as the appraisal frame becomes active, the appraisals become the emotion. Then, as further processing generates more appraisals, these are added to the emotion.

In terms of 4 (emotional experience), the system has some emotional experience but it is incomplete. The system is cognitively aware of its emotional state (the appraisals and the resulting feeling are available in Soar's working memory). Also, the feelings are subjective in the sense that the agent can, in principle, interpret them however it sees fit. While we did not explore this here, there is nothing that prevents cultural knowledge from being added that would allow the agent to generate labels for or other interpretations of the feeling frame the system generates. However, in the current implementations, it has only a trivial physiological system.

For 5 (mind–body interactions), emotions can influence decision making, in that the agent can decide to give up when its emotional state is bad. We evaluate this quantitatively in the context of coherent behavior below. In Marinier (2008), we describe an extension of this system that learns as well. However, we have not yet explored connections to memories, perception, physiology, or a host of other areas that could be influenced by emotion.

The remaining criterion, 1, is whether the agent exhibits emotional behavior. We will explore this quantitatively below.

Picard's list can be extended with additional requirements. First, while we have described how the model works at the micro level, we have not yet demonstrated that it actually produces useful, purposeful behavior. Does it even finish the task? If not, does its emotional state justify the failure? Furthermore, there are several implications that should be explored. For example, if an agent's feelings are determined by the available stimuli, then different environments should lead to different feelings. Additionally, even in environments where the distance to the goal is the same, since Attend takes information about the situation (in a task-independent representation) into account (e.g., Suddenness), different environments should result in different amounts of time to completion. We also claimed in the last section that feelings should impact behavior, both directly and indirectly. Thus, we suggest that there should be a loop: behavior influences feelings, which influence behavior.

We will show results that suggest the model meets the additional requirements described above (summarized here):

(6) The model works and produces useful, purposeful behavior.
(7) Different environments lead to differences in behavior, including:
    (a) Different time courses
    (b) Different feeling profiles
(8) In a given environment where the agent has choices, these choices impact feelings and thus the agent's success.

As discussed earlier, for simplicity, we used a non-human agent in the synthetic Eaters environment. Thus, while we present time course data, these data should not be mapped onto real time for comparison to humans given the simplicity of the Eaters environment, sensors, and effectors.

5.1. Methodology

To evaluate the agent, we used several different mazes in the Eaters domain with a specific goal location in each. In each maze, the distance from the start to the goal was 44 moves (except for the last maze, in which it was impossible to reach the goal). Our aim in designing these mazes was to place the agent in progressively more difficult situations to demonstrate the properties listed above. In the first maze (Fig. 11a), the agent did not have to ever retask to reach the goal, and there were no distracting stimuli; that is, it could not see any walls on its way to the goal. The second maze (Fig. 11b) is exactly the same as the first except that the path to the goal is lined with walls (and hence distractions). Thus, even though there are fewer possible moves, there are just as many Attend opportunities, and they are actually more interesting (hence, distracting). The third maze (Fig. 11c) is very similar to the second, except that there is a kink in the path that requires a brief retasking to maneuver around. This is because the agent has no direct way of making progress when it reaches the kink – if it moves north, it will be further from the goal, and it cannot move east because of the wall. Thus, retasking allows it to temporarily move further from its original goal. The fourth maze (Fig. 11d) contains twists and turns such that four subtasks are required to reach the goal. In the fifth maze (Fig. 11e), it is not possible to reach the goal.

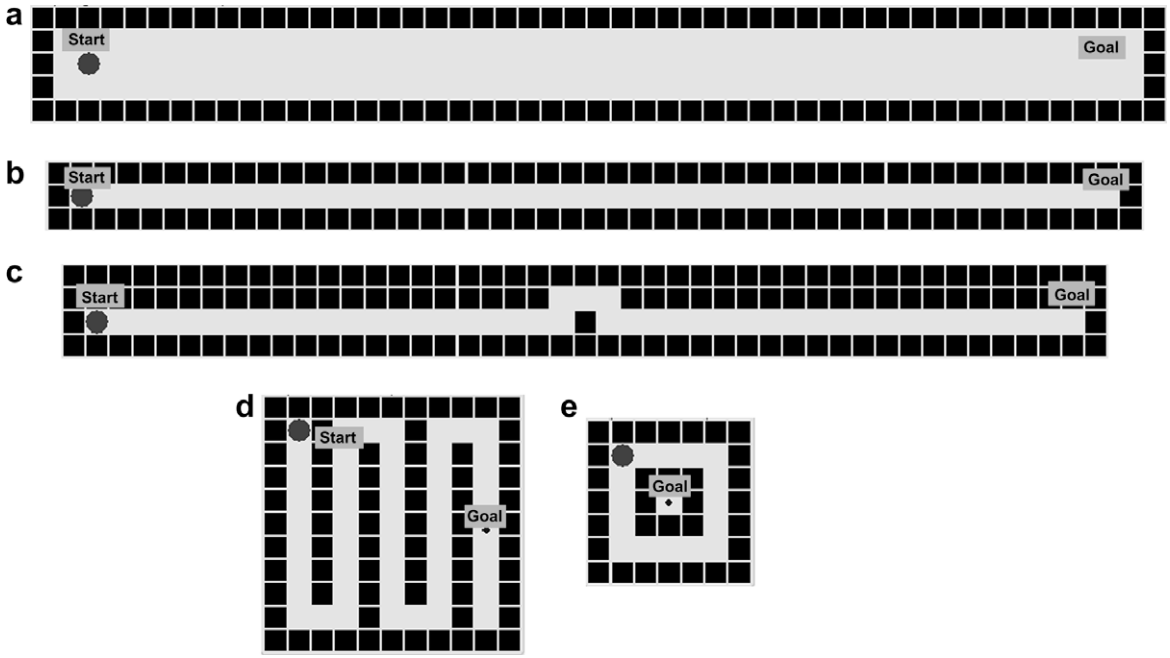

Fig. 11. Eaters mazes. (a) No distractions. (b) With distractions. (c) Distractions and one subtask. (d) Distractions and multiple subtasks. (e) Cannot be completed.

5.1.1. Labeling appraisal frames

While the agent does not use linguistic labels to determine its behavior, we found such a labeling function is useful in analyzing the agent's behavior (indeed, we use it in the results reported here). The labeling function is based on the Manhattan distance between the agent's appraisal frame and the modal emotions defined by Scherer (see Table 1). Since some modal emotions have many unspecified values (which are treated as distance 0), some emotions are frequently closer to the feeling frame than others, even when their specified appraisal values are not good matches. Elation/joy is one such emotion (it has open values for Intrinsic Pleasantness, Discrepancy from Expectation, Control and Power). To compensate for this, we only considered modal emotions that have a Conduciveness with the same sign (or an open Conduciveness). In other words, we divided the emotions into positive and negative emotions based on Conduciveness, and ensured that only labels with the same valence as the frame could be applied. Thus, it is not possible for a feeling with negative Conduciveness to be labeled as elation/joy.

An unusual case in the labeling function is the displeasure–disgust label: Scherer defines it in terms of Intrinsic Pleasantness rather than in terms of Conduciveness (see Table 1), so we split instances of these into positive and negative, as defined by whether Conduciveness was positive or negative. Thus, positive displeasure–disgust is when that label most closely matches the current feeling, but Conduciveness is positive. This can occur when the agent must do something it dislikes, but is necessary to make progress in the task. Real-life examples might be washing the dishes or cleaning a toilet.

5.2. Results

In the first two mazes, the agent will never give up, since it never has to retask. However, we anticipate that the distractions from the walls in the second maze will make it take significantly longer to complete than the first, and that the agent will experience more negative emotions as a result. In the last three mazes, retasking is required and thus the agent can fail. In the third and fourth mazes, the addition of the subtasks require extra processing that could cause the agent to take longer to complete the mazes. Moreover, in the fourth maze, the agent is likely to give up before achieving the goal because of it detects it is not

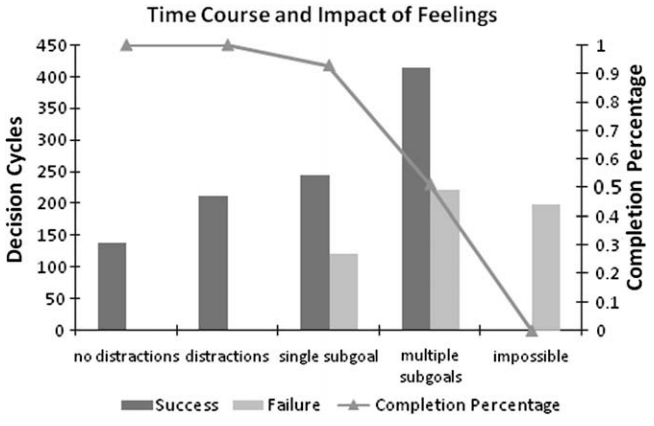

Fig. 12. The bars show the number of decision cycles required to complete each maze for the success and failure cases. The line shows the success rate. All differences are statistically significant (1000 trials for each maze, $> 95\%$ confidence level).

making progress. We expect that the agent will alway give up on the fifth maze because it is impossible to solve. We expect this to take less time than the fourth maze, because in the fourth maze the agent is always making progress, whereas in the fifth maze, after the first subtask, the agent detects that it is not making progress, which should lead the agent to feel worse and hence give up sooner.

Fig. 12 shows the time course of behavior in the different mazes, as well as the success rate in each maze. As we predicted, the mazes do lead to different time courses, which fulfills property 7a (different time courses). In general, as the mazes increase in difficulty, the agent takes longer to complete (or give up on) them. When the agent does give up, though, it takes less time. This makes sense since the agent is stopping early. Still, the maze with multiple subtasks takes longer than the maze with a single subtask when the agent gives up. The impossible maze takes slightly less (but still statistically significant) time to give up. This is because, after the first subtask, all subtasks are considered "bad" subtasks, whereas in the other mazes all subtasks are "good" subtasks. This should mean that there are more negative appraisals in the impossible maze, causing the agent to feel worse and thus give up sooner.

In Fig. 13 we see that the data are consistent with this analysis. The feeling labels in the figure are generated as described in Section 5.1.1. In each maze's feeling profile, the positive feeling (elation–joy) instances outweigh the negative feeling (anxiety–worry and displeasure–disgust) instances except for the impossible maze, where the negative feelings dominate. We can also see that each maze produces a different feeling profile, and that feeling profiles also differ between the success and failure cases. This supports property 7b (different feeling profiles). In contrast, the failure cases for mazes 3 and 4, the positive and negative feelings are nearly equal. This is to be expected given that the subtasks are "good," the agent positively appraises every move it makes (since it thinks it is making progress).

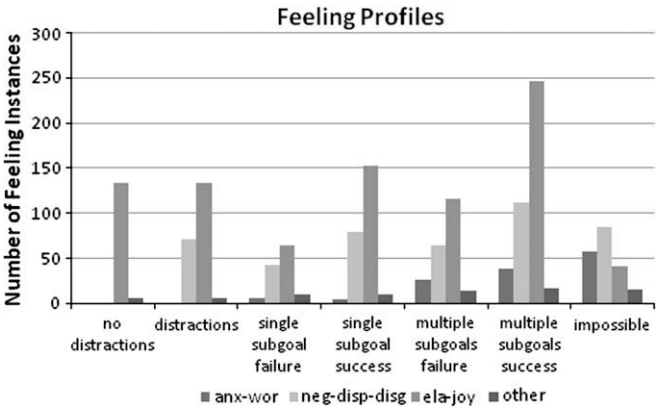

Fig. 13. The average number decision cycles each kind of feeling was active. Labels were produced by our labeling function. Success and failures for mazes 3 and 4 reported separately. "Other" includes Boredom-Indifference, Fear, Positive displeasure–disgust, and Sadness-Dejection. Differences between bars within a group (e.g., no distractions, etc.) are statistically significant (1000 trials for each maze, > 95% confidence level).

Thus, this offsets the negative feelings to some extent. However, each negative feeling in a subtask represents an opportunity to give up, and these more frequent opportunities lead to failure.

This, together with the data from Fig. 12 supports properties 1 (emotional behavior) and 8 (choices influence feelings). That is, success and failure (both absolutely and in terms of rate) are defined by different feeling profiles, implying that feelings do influence behavior. Furthermore, even within the same maze the success and failure cases have different profiles, implying that the choices the agent makes in those mazes impacts feelings and behavior.

Finally, the above analysis supports property 6 (purposeful, useful behavior). That is, the agent's behavior and feeling profiles are expected given its task and environments. The agent completes the task in many cases, and when it fails, it has a negative feeling profile which justifies giving up.

As a final comment, as shown in Fig. 13, the agent experiences a wide breadth of feeling types in these mazes (seven different kinds according to our labeling function). Given the limited nature of the domain, one might expect a much more limited set of feelings. Indeed, we have shown that multiple feelings can arise from simple manipulations of the environment, even in similar situations. One way is via interactions with the goal – adding structure that requires subtasks leads to many different feelings emerging. Another way is via interactions between mood (including decay) and emotion. Sometimes, even though we might classify a mood one way and an emotion another way, their combination results in yet another classification. This prediction could help explain why people are sometimes confused about their feelings.

6. Related work

Like the Soar system we described in this paper, there are several implemented computational systems that use appraisal theory in some form and realize a functional agent that can behave in some environment, and in fact systems such as Gratch and Marsella's (2004) EMA (EMotion and Adaptation) inspired the current work. The primary goal of these systems is generating believable behavior, and there is less of an emphasis on the underlying theoretical integration of emotion and cognition, beyond the assertion that cognition is required to generate appraisals. In addition to different goals, these systems differ from the Soar system in two theoretically important ways. First, most existing systems generate appraisals and emotions all at once and then only rely on the emotion *outcome*. That is, while the emotion has an impact on the system, the appraisals do not. This property can be appreciated by observing that the emotion generation could occur via a non-appraisal process, and the system would not know the difference. In contrast, appraisal generation is required as part of the Soar agent's normal processing – they cannot be replaced by some other emotion-generation process.

Table 4
Comparison summary

System	Appraisal theory	Emotion type	Mood/feeling	Incremental appraisals	Appraisals required
EMA	Mixture	Categorical single	Mood only	Yes	Coping only
MAMID	Mixture	Categorical multiple	No	No	No
OCC/Em	Mixture	Categorical multiple	Mood only	No	No
Kismet	Circumplex	Categorical single	No	No	No
Our system	Scherer	Continuous single	Yes	Yes	Yes

Second, a consequence of appraisals being generated as part of the Soar agent's normal processing is that there is a time course to the generated appraisals and resulting emotions so that the during processing of a single stimulus, the agent's emotions can change as new information becomes available. Many existing systems do not support this because the appraisals are generated all at once.

In the remainder of this section we will briefly describe various systems with respect to these two distinguishing issues, as well as several other dimensions, including system type (architecture or modular), which appraisal theory is used, how many emotions the system can have and whether they are categorical or continuous, and whether it has mood and feeling. Table 4 summarizes the comparison.

EMA is a computational model of a simple appraisal theory implemented in Soar 7 (an older version of Soar). EMA uses its own appraisal theory based on common dimensions from several existing theories. Like our model, appraisals are generated incrementally, but attention does not gate the generation of later appraisals. Rather, EMA generates multiple appraisal frames at once, and an attention mechanism focuses on a single frame, which determines the emotion. One or more categorical labels are then assigned to the single emotion instance; we interpret this as more specific emotion labels, as opposed to multiple emotions. EMA also has mood, which is an aggregate of all current appraisal frames; in contrast, mood in our system is an aggregate over previous emotions (including the current emotion). Finally, the appraisals are required by EMA's coping mechanism, but not directly by other mechanisms (e.g., the attention mechanism uses emotion intensity, but not the appraisals).

MAMID (Hudlicka, 2004) is a system aimed at building emotions into a cognitive architecture. MAMID's architectural mechanisms are higher level than Soar's, making it more a modular system by comparison. For example, it has a Situation Assessment module and an Action Selection module, as opposed building these out of more primitive components. Like EMA, the appraisals used are common to many theories. Unlike our system, MAMID generates an intensity for each of several categorical emotions. While this is modulated by the previous emotion, there is no separate mood concept. Appraisals in MAMID are generated "all at once," in the sense that the Affect Appraiser module takes in information about the current situation and outputs an emotional state. Thus, appraisal is not necessarily required by the system, and could be replaced by some other method for generating emotion.

Ortony et al. (1988) describe a theory (commonly called the OCC model) that was not originally intended for use in systems that have emotion, but has since been implemented for that purpose. We will discuss OCC in the context of Neal Reilly's (1996) Em system. As a theory, OCC does not specify the architecture of the underlying system, but Em is implemented as a modular system. OCC uses a small set of appraisals inspired by existing theories to generate an emotion hierarchy. In Em, multiple categorical emotions can exist simultaneously. OCC only briefly touches on mood, but leaves it unspecified. In Em, mood is an aggregation of current emotions, similar to how EMA uses an aggregate of current appraisal frames. Like MAMID, Em uses an Emotion Generation module that takes a situation description and outputs an emotion – the fact that it uses OCC (and hence appraisal) internally is not critical to its functioning. Like MAMID, then, appraisals are not generated incrementally.

Kismet (Breazeal, 2003) is a social robot. It is a modular system, but as a functioning robot, it handles real perception and motor. It also has physiological drives. While it has "appraisals," these are arousal, valence, and stance, which are better described as a circumplex model (Yik et al., 1999). Kismet can be in a single categorical emotion state at a time, and there is no mood (although the current emotion can indirectly influence the next emotion). Appraisal is not incremental, in the sense that all appraisal dimensions always have a value. Additionally, the appraisal information is only used to generate the emotions, and thus is not actually required by the system.

7. Future work

There are vast, overlapping areas we have yet to explore. One goal is to expand to a more complete model of emotion, including its integration with the rest of cognition and physiology. This expansion will likely provide additional constraints to help shape our theory, and our theory may provide additional constraints on the theories in these areas. For example, how we represent appraisals and emotion may be influenced by these other areas, and vice versa. Besides these areas, we will also discuss scalability, and validation.

Beyond our very abstract mood model, the system has no notion of physiology. Physiology plays critical roles in action tendencies (Frijda et al., 1989), non-verbal communication such as facial expression (Ekman et al., 1987) and tone of voice, and other more basic physiological measures

such as skin conductance, heart rate, and blood pressure. Once a more complete physiological model is in place, we can also explore introspection about the current physiological state, for example, which may extend to emotion recognition (Picard, 1997). Basic drives such as hunger and thirst can also be explored in the context of emotion.

On the cognitive side, we have already scratched the surface of learning elsewhere (Marinier, 2008; Marinier & Laird, 2008), but that remains a major area for continued research. For example, we have not yet explored how appraisal values might be learned. We also need to explore how emotion interacts with other cognitive mechanisms; for example, the episodic and semantic memories depicted in Fig. 2. Such a system could allow phenomena ranging from priming effects (Neumann, 2001) to emotional intelligence (Picard, 1997) to be explored. There is also the major issue of whether the system described here will scale to more complex environments and more complex appraisal value generation (both of which we began to explore in Marinier, 2008). But there is also the matter of simply generating more appraisals; for example, what about socially oriented appraisals? Does the system scale to explaining aspects of social interaction and culture?

Finally, there is the issue of validation. There are multiple ways in which we might attempt to validate the system going forward: believability (Neal Reilly, 1996), human data, including timing (which we briefly explore in Marinier, 2008), physiological measures, behavior, and decision making (Gratch, Marsella, & Mao, 2006), and functionality (e.g., learning, impact of additional appraisals, etc.), which we have started exploring (Marinier, 2008; Marinier & Laird, 2008).

This partial list demonstrates the vast amount of work remaining; it seems unlikely that anything short of a complete human intelligence system can actually address it all. Indeed, this is perhaps a key point that emotion researchers have been making for a long time: emotion is a key aspect of human-level intelligence.

8. Conclusion

We have presented a novel integration of cognition and emotion based on the functional fit between appraisal theory and an abstract theory of cognitive control (PEACTIDM): cognition (as PEACTIDM) provides the processes necessary to generate emotions, whereas emotion (via appraisals) provides the data which cognition (via PEACTIDM) functionally demands. To evaluate the feasibility of this theory, we extended the Soar cognitive architecture to include the computational mechanisms necessary to support our proposed integration. We explored this system within the context of a simple stimulus response task and an ongoing task. Our evaluation centered on qualitative and quantitative issues regarding whether the system actually works and has features consistent with a complete emotion system. For the most part, it succeeds, although we discussed several avenues for future expansion.

We summarize the key theoretical features of our proposal as follows:

(1) Appraisals are a functionally required part of cognitive processing; they cannot be replaced by some other emotion generation theory.
(2) Appraisals provide a task-independent language for control knowledge, although their values can be determined by task-dependent knowledge. Emotion and mood, by virtue of being derived from appraisals, abstract summaries of the current and past states, respectively. Feeling, then, augments the current state representation with knowledge that combines the emotion and mood representations and can influence control.
(3) The integration of appraisal and PEACTIDM implies a partial ordering of appraisal generation.
(4) This partial ordering specifies a time course of appraisal generation, which leads to time courses for emotion, mood and feeling.
(5) Emotion intensity is largely determined by expectations and consequences for the agent; thus, even seemingly mundane tasks can be emotional under the right circumstances.
(6) In general, appraisals may require an arbitrary amount of inference to be generated.

References

Agre, P. (1988). *The dynamic structure of everyday life*. Dissertation, MIT, Electrical Engineering and Computer Science, Cambridge.

Anderson, J. R. (2007). *How can the human mind exist in the physical universe?* New York: Oxford University Press.

Breazeal, C. (2003). Function meets style: Insights from emotion theory applied to HRI. *IEEE Transactions in Systems, Man, and Cybernetics, Part C, 34*(2), 187–194.

Damasio, A. (1994). *Descarte's error: Emotion, reason, and the human brain*. New York: Avon Books.

Damasio, A. (2003). *Looking for Spinoza: Joy sorrow and the feeling brain*. USA: Harcourt.

Ekman, P., Friesen, W., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712–717.

Forgas, J. P. (1999). Network theories and beyond. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 591–611). Chichester, England: Wiley and Sons.

Frijda, N. H., Kuipers, P., & ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology, 57*, 212–228.

Gratch, J., Marsella, S., & Mao, W. (2006). Towards a validated model of "Emotional Intelligence". In *21st national conference on artificial intelligence*. Boston: AAAI.

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

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*, 348–362.

Hudlicka, E. (2004). Beyond Cognition: Modeling Emotion in Cognitive Architectures. In Proceedings of the international conference of cognitive modelling, ICCM 2004 (pp. 118–123). Pittsburgh, PA.

Kieras, D., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human–computer interaction. *Human–Computer Interaction, 12*, 391–438.

Laird, J. (2008). Extending the Soar cognitive architecture. In *Proceedings of the first conference on artificial general intelligence*. Memphis: IOS Press.

Marinier, R. (2008). *A computational unification of cognitive control, emotion, and learning*. Dissertation, University of Michigan, Ann Arbor.

Marinier, R., & Laird, J. (2007). *Computational modeling of mood and feeling from emotion. CogSci 2007*. Nashville: Cognitive Science Society.

Marinier, R., & Laird, J. (2008). *Emotion-driven reinforcement learning. CogSci 2008*. Washington, D.C.: Cognitive Science Society.

Marsella, M., & Gratch, J. (2009). EMA: A process model of appraisal dynamics. *Cognitive Systems Research, 10*, 70–90.

Neal Reilly, W. S. (1996). *Believable social and emotional agents*. Technical Report CMU-CS-96-138, Carnegie Mellon University, Pittsburgh.

Neal Reilly, W. S. (2006). Modeling what happens between emotional antecedents and emotional consequents. In *Proceedings of the eighteenth European meeting on cybernetics and systems research* (pp. 607–612). Vienna, Austria: Austrian Society for Cybernetic Studies.

Neumann, R. (2001). The causal influences of attributions on emotions: A procedural priming approach. *Psychological Science, 11*(3), 179–182.

Newell, A. (1990). *Unified theories of cognition*. Cambridge: Harvard University Press.

Newell, A., Shaw, J. C., & Simon, H. A. (1960). Report on a general problem solving program. In *Proceedings of the international conference on information processing* (pp. 256–264). Paris: UNESCO.

Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge, MA: Cambridge University Press.

Parkinson, B. (2009). What holds emotions together? Meaning and response co-ordination. *Cognitive Systems Research, 10*, 31–47.

Picard, R. (1997). *Affective computing*. Cambridge, MA: MIT Press.

Reisenzein, R. (2009). Emotions as metarepresentational states of mind: Naturalizing the belief-desire theory of emotion. *Cognitive Systems Research, 10*, 6–20.

Roseman, I., & Smith, C. A. (2001). Appraisal theory: Overview, assumptions, varieties. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research*. New York: Oxford University Press.

Rosenberg, E. L. (1998). Levels of analysis and the organization of affect. *Review of General Psychology, 2*, 247–270.

Scherer, K. (2001). Appraisal considered as a process of multilevel sequential checking. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research*. New York: Oxford University Press.

Schorr, A. (2001). Appraisal: The evolution of an idea. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 20–34). New York: Oxford University Press.

Smith, C. A., & Kirby, L. A. (2001). Toward delivering on the promise of appraisal theory. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory methods, research* (pp. 121–138). New York: Oxford University Press.

Smith, C. A., & Lazarus, R. S. (1990). Emotion and adaptation. In L. A. Pervin (Ed.), *Handbook of personality theory and research* (pp. 609–637). New York: Guilford.

Sun, R. (2006). The CLARION cognitive architecture: Extending cognitive modeling to social simulation. In R. Sun (Ed.), *Cognition and multi-agent interaction*. New York: Cambridge University Press.

Yik, M., Russell, J., & Feldman Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology, 77*(3), 600–619.

Zajonc, R. (1984). On the primacy of affect. *American Psychologist, 39*, 117–123.