

Modeling Interaction Between Metacognition and Emotion in a Cognitive Architecture

Eva Hudlicka

Psychometrix Associates, Inc.
1805 Azalea Drive, Blacksburg, VA , USA
evahud@earthlink.net

Abstract

While research in metacognition has grown significantly in the past 10 years, there has been a relative lack of research devoted to the focused study of the interactions between metacognition and affective processes. Computational models represent a useful tool which can help remedy this situation by constructing causal models of demonstrated correlational relationships, and by generating empirical hypotheses which can be verified experimentally. In this paper we describe enhancements to an existing cognitive-affective architecture that will enable it to perform a subset of metacognitive functions. We focus on modeling the role of a specific metacognitive factor, the *feeling of confidence (FOC)*, and the anxiety-linked metacognitive strategy of emotion-focused coping.

Introduction and Objectives

Metacognition, defined variously as ‘thoughts about thoughts’ or ‘awareness and control’ of one’s thoughts, is considered by many to be an essential component of skilled performance, influencing memory functions, learning and skill acquisition, and problem-solving. While research in metacognition has grown significantly in the past 10 years, there has been a relative paucity of research devoted to the focused study of the interactions between metacognition and affective processes. A notable exception is the work of Wells (2000), and Matthews and Wells (2004). However, this work focuses on metacognition and emotion in the context of emotional disorders (e.g., depression, generalized anxiety, obsessive compulsive disorders), rather than on the role of emotion in normal subjects, or the interaction between emotion and metacognition in transient affective states (e.g., high stress or frustration).

Computational models represent a useful tool for elucidating the mechanisms of cognitive processing. Ideally, these models are based on existing empirical data and used to confirm hypothesized mechanisms. However,

in situations where such data are lacking, these models can be helpful in generating empirical hypotheses for further experimental testing and data gathering.

In this paper we describe enhancements to an existing cognitive-affective architecture (MAMID) (Hudlicka, 2002; 2003), which support the exploration of affective-metacognitive interactions. The enhancements will enable MAMID to perform a subset of the metacognitive functions involved in monitoring and control of cognition, and the associated metacognitive knowledge and beliefs. The initial focus will be on modeling the role of a specific metacognitive factor, the *feeling of confidence (FOC)*, and the *anxiety-linked metacognitive strategy of emotion-focused coping*.

The intended benefits of the model are in both the theoretical and the applied realms. In the *theoretical realm*, the exercise of building a model requires an operationalization of concepts and relationships which help refine existing psychological theories, and generate empirical hypotheses for further testing. In the *applied realm*, the explicit model of metacognition, and its interactions with affective factors, promises to provide a more realistic model of human behavior, both adaptive and maladaptive (e.g., models where metacognition diminishes performance), and generate more effective agent behavior (e.g., improved performance under stress).

This paper is organized as follows. First, we briefly summarize key findings about metacognition and about known interactions between metacognition and emotion. Next, we describe the existing MAMID cognitive-affective architecture, within which the proposed metacognitive component will be implemented. We then describe the proposed design for modeling the role of FOC in metacognitive monitoring and control, and the metacognitive knowledge required. We then outline two models where emotion influences metacognitive processing. We conclude with a brief discussion of related work, a summary and future work.

Metacognition and Its Interaction with Emotion

What is metacognition? The simplest definition of metacognition is “thinking about thinking” (Nelson, 2002). However, the simplicity of this statement belies the complexity and diversity of processes and structures that mediate the variety of identified metacognitive activities. More differentiated definitions distinguish between awareness (and associated monitoring functions of cognition), and control (and associated executive and self-regulatory functions of cognition) (Osman and Hannafin, 1992; Nelson and Narens, 1990). A more encompassing definition states that metacognition is a “multifaceted concept comprising the knowledge and beliefs, processes and strategies that appraise, monitor or control cognition” (Wells, 2000). Metacognitive knowledge is then defined as knowledge individuals have about their own cognitions, as well as about the factors that influence their cognitions.

What is the role of metacognition? Evidence indicates that metacognitive control and regulation is comprised of a range of functions including attention allocation, checking, planning, memory retrieval and encoding strategies, and detection of performance errors (Wells, 2000). In general, metacognition is involved in strategy selection for complex problems requiring resource tradeoffs, for dealing with unfamiliar situations, and for troubleshooting. A number of researchers discuss the fact that metacognition can be helpful, neutral, or harmful to cognition and performance (e.g., Paris, 2002).

Relationship Between Metacognition and Emotion As stated above, data regarding the mutual influences among emotion and metacognition are unfortunately limited, and focused almost exclusively on psychopathology (e.g., Wells, 2000; Matthews and Wells, 2004). For the purpose of modeling, we need to identify the specific effects of particular affective factors (states or traits) on particular metacognitive functions and knowledge. To help organize the known effects, and to identify gaps in data, it is useful to categorize the effects into those resulting from states vs. traits, and those affecting processing mechanisms vs. knowledge structures. Examples of identified correlations include: *State effects on processes*: Anxiety-linked appraisal of events as threats; emotion-focused coping; Depression-linked self-criticism focused coping; *Trait effects on processes*: Neuroticism-linked preference for self-information; *Trait effects on knowledge*: Neuroticism-linked predominance of negative schemas

(threat, negative self evaluations, negative future projections).

MAMID Cognitive-Affective Architecture and Modeling Methodology

Here we briefly describe the existing cognitive-affective architecture which will be enhanced with the proposed metacognitive functions. A key component of the architecture is an affect appraisal module, which dynamically generates affective states as a function of both internal and external factors (e.g., incoming cues, internal situation assessments and goals), and both dynamic and static agent attributes (e.g., prior existing emotion, stable personality trait profile). We also discuss the generic modeling methodology used to model the interacting effects of states, traits and other individual differences in terms of parametric manipulations of the architecture processes and structures.

MAMID Cognitive Architecture The cognitive architecture implements a sequential see-think-do processing sequence (figure 1), consisting of the following modules: *sensory pre-processing*, translating incoming data into task-relevant cues; *attention*, filtering incoming cues and selecting a subset for processing; *situation assessment*, integrating individual cues into an overall situation assessment; *expectation generation*, projecting current situation onto possible future states; *affect appraiser*, deriving the affective state (both valence and four of the basic emotions) from a variety of external and internal elicitors, both static and dynamic; *goal selection*, selecting critical goals for achievement; and *action selection*, selecting the best actions for goal achievement.

These *modules* map the incoming stimuli (cues) onto the outgoing behavior (actions), via a series of intermediate internal representational structures (situations, expectations, and goals), collectively termed *mental constructs*. This mapping is enabled by long-term memories (LTM) associated with each module, represented in terms of belief nets or rules. *Mental constructs* are characterized by their attributes (e.g., familiarity, novelty, salience, threat level, valence, etc.), which influence their processing; that is, their rank and the consequent likelihood of being processed within a given architecture cycle; (e.g., cue will be attended, situation derived, goal or action selected). (Note that the availability of the mental constructs from previous frames of the execution cycle allows for dynamic feedback among constructs, and thus departs from a strictly sequential processing sequence.)

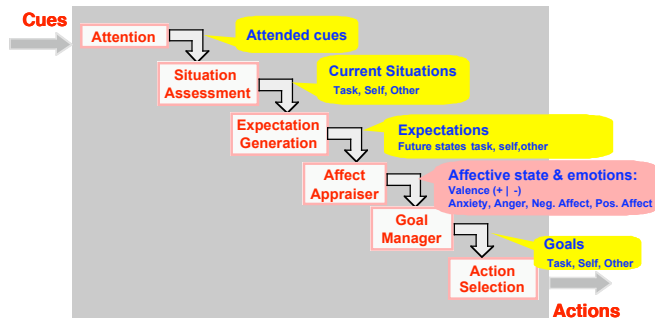


Figure 1: MAMID Cognitive Architecture: Modules & Mental Constructs

The Affect Appraisal module is a core component of the MAMID architecture. It integrates external data (cues), internal interpretations (situations, expectation) and priorities (goals), and stable and transient individual characteristics (traits and existing emotional states), and generates an affective appraisal in terms of both a valence (positive / negative) and one of the four basic emotions (fear/anxiety, anger/frustration, sadness, joy).

The basic emotions are calculated via a series of belief nets stored in the agent's LTM. Differences in the triggering elicitors for particular emotions allow for the representation of individual idiosyncracies in emotion triggering (e.g., Agent A might react to situation x with anger, Agent B with fear, whereas Agent C might not have an affective reaction at all.) The model incorporates elements of several recent appraisal theories: *multiple-levels* and *multiple stages* (Leventhal & Scherer, 1987; Smith & Kirby, 2001).

Generic State / Trait Modeling Methodology To model the interacting effects of traits and states on cognitive processing, MAMID uses a previously described methodology (Hudlicka, 2002; 1998), which consists of mapping particular state / trait profiles onto specific architecture parameter values (figure 2). These parameters then control processing within individual architecture modules.

Functions implementing these mappings were constructed on the basis of the available empirical data. For example, reduced attentional and working memory (WM) capacity, associated with anxiety and fear, are modeled by dynamically reducing the attentional and WM capacity of the architecture modules, which then reduces the number of constructs processed (fewer cues attended, situations derived, expectations generated, etc.). Attentional threat bias is modeled by higher ranking of threatening cues, thus increasing their likelihood of being attended, and by higher ranking of threatening situations and expectations, thus increasing the chances of a threatening situation / expectation being derived. Trait-linked structural differences in LTM are supported by allowing the flexible selection of alternative LTM clusters, reflecting distinct personality traits. Traits also

influence the dynamic characteristics of the emotional responses (ramp up, decay, and maximum intensities).

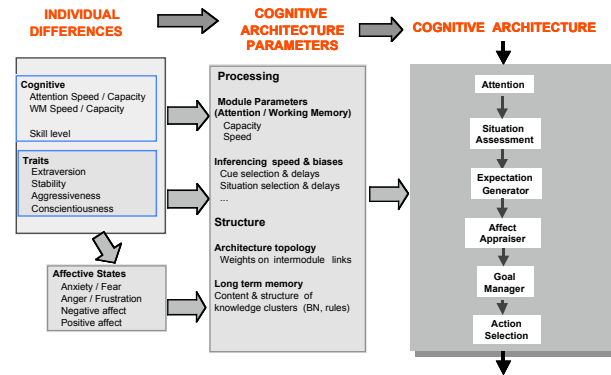


Figure 2: Parametric State / Trait Modeling Methodology

The initial version of MAMID was implemented in the context of a peacekeeping scenario, with each instance of the MAMID architecture controlling the behavior of a simulated Army commander, reacting to a series of surprise situations (e.g., ambush, hostile crowd) (Hudlicka, 2003). MAMID was able to demonstrate distinct processing differences due to the different trait profiles and dynamically generated states, with the distinct commanders behaving differently during the course of the scenario, leading to differences in mission outcomes. The domain-independent MAMID architecture is currently being transitioned into a NASA context, and it is within this context that the metacognitive enhancements described below will be implemented.

Proposed MAMID Enhancements Implementing Metacognitive Functions

The objective of the enhancements described below is to augment the existing MAMID architecture with the ability to perform a subset of metacognitive functions, and to explicitly model interactions among selected metacognitive functions and emotion. Below we describe a model of the *feeling of confidence* factor, and its role in triggering metacognitive control strategies. In the next section we describe two examples focusing on emotion.

Feeling of Confidence (FOC) FOC is a component of metacognition that reflects the level of confidence in particular cognitions. FOC judgments can refer to past, present and future cognitive activities, and apply to a variety of cognitive processes, including memory retrieval, problem-solving, and planning. FOC is thought to play a role in controlling cognitive iteration during problem-solving and memory retrieval, by determining whether a particular answer will be accepted, or whether

further cognitive refinement will be necessary (e.g., Narens et al., 1994).

To model FOC and its role within MAMID we introduce a meta-layer “above” the current MAMID processing sequence (refer to figures 1 and 3). The meta-layer contains two modules, “Monitoring” and “Control”, each of which accesses the metacognitive knowledge-base. Similarly to its object-level analogs, the metacognitive knowledge base consists of rules and belief nets which contain beliefs and knowledge about cognitions, and rules for selecting particular metacognitive monitoring, control, and coping strategies.

To model FOC and its role, two other features are introduced to the existing MAMID model. *First*, each mental construct is augmented to include an FOC attribute. Thus a distinct FOC is associated with all of the generated *cues, situations, expectations, goals and actions*. This FOC value reflects the agent’s confidence that the corresponding constructs represent the actual stimuli (wrt cues), accurate interpretations (wrt situations), accurate projections (wrt expectations), appropriate priorities (wrt goals), and appropriate responses (wrt actions). Initially, FOC will be calculated using a combination cognitive and affective factors, which will include: anxiety (reducing FOC), awareness of alternatives (inversely proportional to FOC), task difficulty (inversely proportional to FOC), and awareness of lack of knowledge (inversely proportional to FOC) (Efklides, 2002).

Second, a distinct FOC threshold is introduced for each construct *type*. Thus there is a situation FOC threshold, expectation FOC threshold, action FOC threshold etc.

When a mental construct is produced by the corresponding module during the processing sequence at the object level, the construct’s FOC is compared to the corresponding construct-type FOC threshold (e.g., FOC (situation x) is compared with the FOC threshold for the construct type ‘situation’, etc.). If the construct FOC is *greater than* the corresponding construct type threshold, then the feeling of confidence is adequate and no further metacognitive activity is required. If the FOC is *lower than* the FOC threshold, this signals that the desired level of subjective confidence has not been reached and metacognitive *control* activity is triggered, in an attempt to increase FOC, or to shift strategies entirely.

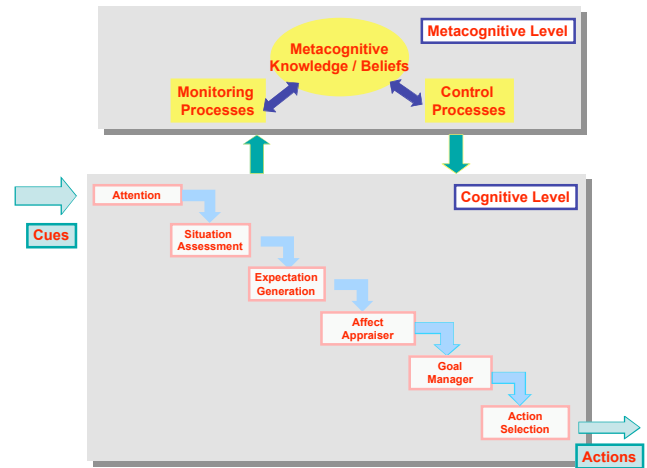


Figure 3: Augmenting MAMID With a Metacognitive Layer

What exact form the metacognitive control then takes depends on the construct type, the contents of the metacognitive knowledge-base, and both the situational context (external factors) and the agent’s internal context (currently activated constructs and emotional states).

Several options exist. One option of course is to do nothing and continue processing at the object level, with the lower-than-desired feeling of confidence. (There may be interesting cumulative effects of this on further future processing, such as an increased sense of subjective uncertainty which may contribute to stress which in turn has a series of specific effects on attention, perception and cognition.) Another option is to re-derive the construct in an attempt to increase the feeling of confidence. The nature of this re-derivation process depends on several factors, including the *position in the processing sequence* of the construct in question, and the type of re-processing possible given the current *informational context*. Both are explained below.

Position in processing sequence The further along the processing sequence a construct is, that is, the closer it is to the action (end) as opposed to the cue (beginning), the more extensive the re-processing can be. To re-calculate the FOC for a cue, the only option is for the attention module to re-scan and obtain a new set of cues. In contrast, to recalculate an action FOC, the agent has the option to re-calculate the action FOC within the action selection module, and to go back recursively through the preceding modules, recalculating the FOC’s of the intervening constructs, and perhaps including an attentional re-scan to obtain new cues.

Informational context By informational context we mean the nature of the data available for (re)processing. For example, to recalculate the situation certainty during a

‘second pass’ through the situation assessment module, the agent has several options: (1) using different subset of the existing information to re-calculate the situation and its associated FOC (e.g., use different schemas, cues – perhaps pursuing options which were rejected in the previous pass); (2) using a different weighting scheme for the existing information (e.g., using the same cues and schemas to calculate the information but changing the relative importance of the constituent cues in contributing to the final situation, assigning more or less weight to self-related cues); or (3) attempting to obtain additional information by directing the attention module to collect additional cues, which may then contribute to a change in the situation FOC.

The choice among these alternatives depends on the current dynamic context (e.g., the current configuration of mental constructs and the agent’s affective state), in conjunction with the contents of the metacognitive long-term memory (mLTM), which stores the particular metacognitive control strategies and their specific triggers.

Thus, for example, a particular agent’s mLTM may contain rules that indicate that when a situation FOC is below the desired threshold, the metacognitive procedure is to re-scan the environment, obtain an update cue set, and recalculate the situation using the standard procedure. (This process also enables the modeling of the familiar confirmation bias, where cues are selectively attended which confirm an existing interpretation and reject alternative interpretations.) Another agent may have an alternative strategy that involves a repetition of the situation assessment process, incorporating “weak” cues that had been rejected in the previous pass, in an attempt to increase the FOC.

A number of possible options can be explored here, including specific content-related variations in strategies, both across and within agents. For example, within the same agent, different control strategies might be used for situations involving the self, a particular task or another specific individual. Different strategies may also be linked to different affective states. For example, during a low state of anxiety, low action FOC may trigger the re-calculation of the action FOC using different data (e.g., taking into consideration a broader range of triggering situations and expectations, in addition to the goal). In contrast, during a high state of anxiety, a low action FOC may trigger an immediate re-scan of the environment in an attempt to restart the processing sequence, in the hopes that a higher action FOC will be achieved.

The contents of the mLTM that define these alternative strategies are specified by the modeler and, to the extent possible, based on available empirical evidence. When the necessary data are lacking, the modeler must make assumptions and generalize from available data. The resulting model can then be evaluated according to the appropriate relevant criteria (e.g., “Does it generate useful

empirical hypotheses which can be verified experimentally?” “Does the agent produce more realistic and effective behavior in simulated situations as a result of the metacognitive enhancements?” etc.), and the mLTM contents are then iteratively refined until the desired performance criteria are obtained.

Modeling Interactions Among Metacognition and Emotions

Below we describe how the proposed design integrates the metacognitive functions with the existing affective components of the MAMID architecture. We focus on models of two specific phenomena: one supported by existing empirical data (anxiety-related emotion-focused coping), and the other a speculative model of the possible relationship between affective factors and the FOC discussed above.

Anxiety-related emotion-focused coping Empirical data indicate that particular state and trait values are linked to distinct styles of coping. Specifically, heightened states of anxiety, and a high score on the neuroticism trait, correlate with emotion-focused coping strategies (Wells, 2000). These are problem-solving strategies that focus on altering an undesirable emotion state, rather than altering the features of the environment that gave rise to that state. Examples of emotion-focused coping strategies include attempts to calm self by distraction, excessive worry, and avoidance. In many cases, this choice of coping exacerbates the problem, since it may not effectively eliminate the original source of the negative emotion.

MAMID already implements this capability, by a combination of several mechanisms: (1) by the dynamic calculation of affective states; (2) by the ability of a particular state-value pair to trigger the selection of a particular goal or a specific action (e.g., IF (anxiety = high) THEN (communicate with peer); IF (anxiety = high) THEN (goal = ‘reduce anxiety’), etc.); and (3) by making a distinction between self- and task-related mental constructs, which allows the preferential processing of one or the other type of construct (this type of bias is linked to several state – trait combinations).

In the proposed enhanced model, we will augment the set of possible coping strategies, and will explicitly represent hierarchies of these strategies, within the mLTM knowledge-base. Thus the mLTM will contain rules linking specific emotions, and emotion-trait combinations to either problem-focused or emotion-focused coping strategies. Within these broad categories, further refinements will then be possible, allowing choices among a broader range of both problem-focused (e.g., improved planning, focus on removal of negative stimulus, finding help, etc.), and emotion-focused strategies (acceptance, venting, avoidance, etc.). This

richer repertoire of strategies will support more realistic agent behavior, and provide the means for exploring models of how alternative strategies are selected and what role metacognition plays in this process.

Speculative model of FOC-affect interactions In this example we describe how a model of the FOC and its interactions with emotions can help elucidate the mechanisms underlying a particular affective and behavioral disorder: obsessive-compulsive disorder. One characteristic of this disorder is excessive checking behavior. This may be due to an abnormally high FOC threshold, which demands repeated “checks” of the performed action, to assure that it has been done. A more internally-oriented characteristic may result in excessive planning and re-planning without ever taking an action – the familiar ‘paralysis by analysis’ phenomenon. While the exact role of FOC in these phenomena has not been empirically established, the roles outlined above represent plausible hypotheses. Below we describe how they would be modeled in the enhanced MAMID architecture.

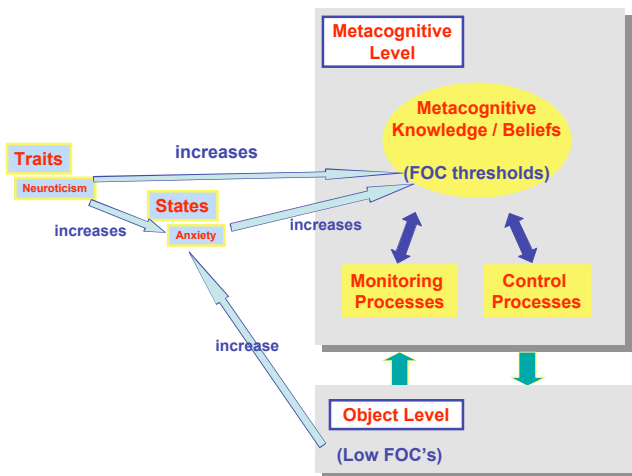


Figure 4: Relationships Between States, Traits, and the Feeling of Confidence Mechanism

Existing data suggest that obsessiveness correlates with both high degree of consciousness and a high state of anxiety (Matthews & Deary, 1998). To model effects of obsessiveness on FOC, MAMID will use the value of the consciousness trait and the level of anxiety to calculate the FOC thresholds for the mental constructs. Since the empirical data are lacking to determine what the exact effects of the contributing factors are, we begin with a first approximation function that captures the qualitative relationship of these two factors to the mental construct thresholds, by having each factor contribute .5 to the final FOC threshold, whose normalized value range is [0-1].

Inclusion of this relationship in the model explicitly links the affective state into the FOC-triggered metacognitive-cognitive processing feedback cycle. It

allows the modeling of both adaptive and maladaptive sequences.

Adaptive Sequence Low FOC values for a particular mental construct trigger anxiety which raises FOC threshold. The FOC construct / threshold discrepancy triggers metacognitive processing, which attempts to increase the construct FOC. This is successful, which leads to a reduction in the level of anxiety which in turn reduces the FOC threshold. Metacognitive activity intervened temporarily to correct the problem.

Maladaptive Sequence In an obsessive-compulsive disorder, this regulatory feedback system is disrupted. The high level of anxiety, coupled with inadequate coping strategies, prevents the derivation of adequately high FOC values, this perpetuates the high level of anxiety, which maintains the high FOC threshold. As a result, the agent is unable to arrive at a decision and remains ‘stuck’ in internal processing and re-processing of existing information.

Related Work

While to our knowledge there is no computational modeling effort focusing exclusively on modeling interaction among metacognition and emotion, there is a large number of systems that use metacognitive functions (reasoning about reasoning, metalevel-control) to improve agent performance, to explore human learning and human metacognition (e.g., Reber & Schunn, 1996), to explore architectural requirements for adaptive and intelligent behavior (e.g., Sloman, 2002), or to improve particular type of processing (e.g., natural language processing, planning, or aspects of HCI). There are at this point also numerous computational models of emotion, and some models of coping (e.g., Gratch and Marsella, 2004). There are also important models in psychology, which have not yet been implemented. An example most relevant to the current effort is Wells’ and Matthews S-REF model (Wells, 2000; Matthews and Wells, 2004).

Summary, Conclusions an Future Work

We described an existing cognitive-affective architecture and the design extensions that would enable it to explicitly model selected metacognitive functions, and their interaction with several affective factors. We focused on a model of the feeling of confidence factor, its role in triggering metacognitive processing, and the metacognitive control alternatives available for re-processing to achieve a greater FOC. We then described a speculative model of possible interactions between the FOC and affective factors. We also described how the design would model the known anxiety-linked emotion-focused coping.

The near-term future work will focus on implementing the proposed metacognitive component within the context

of the existing MAMID architecture. We will then evaluate the resulting model in terms of two criteria: realism and plausibility of the associated agents behavior, and ability to generate empirical hypotheses regarding specific causal mechanisms for empirical exploration of metacognition-emotion interactions.

References

- Cary, M. & Reder, L.M. (2002). Metacognition in Strategy Selection. In Chambres, P., Izaute, M. & Marescaux, P.-J. *Metacognition: Process, Function and Use*. Boston: Kluwer.
- Hudlicka, E. (2002). This time with feeling: Integrated Model of Trait and State Effects on Cognition and Behavior. *Applied Artificial Intelligence*, 16:1-31.
- Hudlicka, E. (2003a). Modeling Effects of Behavior Moderators on Performance: Evaluation of the MAMID Methodology and Architecture, In *Proc. of BRIMS-12*.
- Leventhal, H. and Scherer, K.R. (1987). The relationship of emotion to cognition. *Cognition and Emotion*, 1, 3-28.
- Matthews, G., & Wells, A. (2004) Rumination, depression, and metacognition: The S-REF model. In C. Papageorgiou & A. Wells (Eds.), *Rumination: Nature, theory, and treatment*. Chichester: Wiley
- Mazzoni, G. and Nelson, T.O. (1998). *Metacognition and Cognitive Neuropsychology*. Mahwah, NJ: LEA.
- Metcalf, J. and Shimamura, P. (1994). *Metacognition: Knowing about Knowing*. Cambridge, MA: MIT Press.
- Narens, J., Jameson, K.A. & Lee, V.A. (1994). Subthreshold Priming and Memory Monitoring. In Metcalfe, J. and Shimamura, P. (eds). *Metacognition: Knowing about Knowing*. Cambridge, MA: MIT Press.
- Nelson, T.O. and Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G.H. Bower (Ed.), *The Psychology of learning and motivation*, 26. San Diego, CA: Academic Press.
- Nelson, T.O. (2002). New Theory and Data on Metacognitive Monitoring and Control in Different Contexts and by Different Individuals. In Chambres, P., Izaute, M. and Marescaux, P.-J.. *Metacognition: Process, Function and Use*. Boston: Kluwer.
- Paris, S.G. (2002). When is Metacognition Helpful, Debilitating, or Benign? In Chambres, P., Izaute, M. and Marescaux, P.-J.. *Metacognition: Process, Function and Use*. Boston: Kluwer.
- Reder, L. and Schunn, C. (1996). Metacognition does not imply awareness. In L.M. Reder (ed.), *Implicit Memory and Metacognition*. Mahwah, NJ: LEA.
- Smith, C.A. and Kirby, L.D. (2001). Toward Delivering on the Promise of Appraisal Theory. In *Appraisal Processes in Emotion*. K.R. Scherer, A.Schorr, T. Johnstone (Eds.). NY: Oxford.
- Wells, A. (2000). *Emotional Disorders and Metacognition*. NY: Wiley.