

## ICCM Symposium on Cognitive Modeling of Processes “Beyond Rational”

**William G. Kennedy (wkennedy.gmu.edu)**

Krasnow Institute for Advanced Study, Center for Social Complexity  
George Mason University, Fairfax, VA USA

**Frank E. Ritter (frank.ritter@psu.edu)**

The College of Information Sciences and Technology  
The Pennsylvania State University, University Park, PA 16802 USA

**Christian Lebiere (cl@cmu.edu), Ion Juvina (ijuvina@cmu.edu), Alessandro Oltramari (aoltrama@andrew.cmu.edu)**

Department of Psychology  
Carnegie Mellon University, Pittsburgh, PA 15213 USA

**Jonathan Gratch (gratch@ict.usc.edu)**

Institute for Creative Technologies and Department of Computer Science  
University of Southern California, Playa Vista, CA USA

**Richard M. Young (r.m.young@acm.org)**

UCL Interaction Centre  
University College London, UK

### Abstract

Computational cognitive modeling is normally thought of as rational cognition. However, there are human behaviors that do not appear to be driven by rational cognition. The other, “beyond rational” cognition is also appropriate for computational models of cognition. The panel will discuss their efforts at modeling this form of cognition.

**Keywords:** cognitive models; cognition; Dual Process Theory, emotion, intuition.

### Introduction

Cognitive modeling has been primarily aimed at implementing and testing theories explaining behavior driven by rational, multi-step cognition and it has been very successful (e.g., Anderson, 2007; Anderson, et al., 2004; Laird, 2008; 2012). However, there are many human behaviors that seem to be driven by aspects of behavior that are not the same as “rational” cognition: immediate judgments, intuitive, emotional, and other non-rational, hence “beyond rational” processes. These aspects may result in such phenomena as emotional natural language generation, optical illusions, snap judgments, and humor.

There has been a growing literature on these processes. Significant books include LeDoux’s *The Emotional Brain*, Gigerenzer’s *Gut Feelings*, Klein’s *Sources of Power*, Thagard’s *Hot Thought*, Minsky’s *The Emotion Machine*, and Irvine’s *On Desire*. Herbert Simon also addressed this topic in his *Reason in Human Affairs*. However, computational models of non-rational cognition are relatively rare (cf., Gratch & Marsella, 2004; Kennedy & Bugajska, 2010).

The Dual Process Theory could provide a basis for computational cognitive modeling of these aspects. The

Dual Process Theory suggests two types of processes drive behavior (Evans, 2008; Sloman, 1996). Reality may be more nuanced than a simple dichotomy and this grouping is somewhat controversial. The more neutral terms for the two processes are System 1 and System 2, with System 2 being the rational, conscious, multi-step, slower, more evolutionarily advanced process (Kahneman, 2003). The implicit learning discussion of a few years ago could provide examples of one or the other side, rather than trying to fit all implicit learning phenomena within one side (Wallach & Lebiere, 2002). It may also be that rather than two processes, there may be a spectrum of processes between two extremes or cognition may have more dimensions than one. There is a suggestion that much of our behavior is the result of this other reasoning.

This panel will address the topics related to cognitive modeling of beyond-rational cognition. The panel members will present their views on the topic and whether it would be appropriate for the cognitive modeling community to entertain models of behavior driven by beyond rational processes.

### Panel Makeup

The panel consists of cognitive modelers who have thought about this topic. Each has provided an abstract of their input to this topic.

#### William G. Kennedy

Starting with the ancient Greeks, we have believed that there were two forms of cognition that control our behavior: passion and reason, and that there was an inner battle for control of the mind (LeDoux, 1996). Dualism, proposed by Descartes, separated mind and body and has been

discredited in current philosophy (Evans, 2010). When we began to study cognition scientifically, William James considered reasoning, consciousness, emotion, instinct, and will as separate topics, although consciousness received the shortest treatment (James, 1892/2001).

With the cognitive revolution of the second half of the 20<sup>th</sup> century came a focus on testable theories of Cognitive Science and the verbal descriptions of non-rational cognition have been marginalized. However, recently there has been resurgence in interest in the other side, the intuitive, emotional side of cognition. There have been many books written on how people make decisions using methods outside traditional rational cognition. Dualism has evolved through a dual representation of knowledge, visual and verbal (Paivio, 1971) into a Dual Process theory of cognition (Evans, 2008; Sloman, 1996).

The Dual Process theory suggests a distinct separation of cognitive processes and they can be organized into (at least) four groupings: consciousness, evolution, functional characteristics, and individual differences (Evans, 2008). For example, Table 1 presents the functional characteristics of the two systems.

Table 1: Functional Characteristics of the Dual Process Theory (from Evans 2008).

System 1	System 2
Associative	Rule-based
Domain specific	Domain general
Contextualized	Abstract
Parallel	Sequential

In addition to the concept and the separation of characteristics, even the naming of the two systems is controversial and calling them System 1 and System 2 is an attempt to keep the discussion focused on the content, not the naming (Gray, 2007; Kahneman, 2003).

As an indication of the trend in the interest in the topic, Figure 1 is offered. The figure shows the frequency of searches for the term “System 1 System 2”

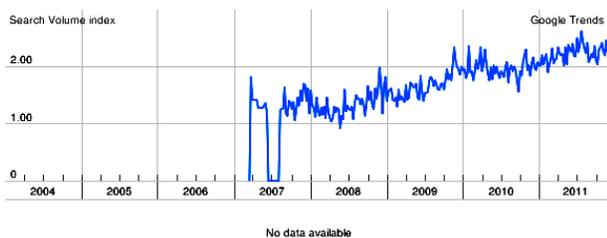


Figure 1. Google Trending of “System 1 System 2”

With the general attention to the concept of dual processes of cognition, the cognitive modeling community should consider more seriously modeling cognition “beyond rational.”

## Frank Ritter

Frank will discuss work on modeling the effects of caffeine on behavior, and work on modeling the effects of stress on behavior. These approaches have been done with sets of changes overlaying the cognitive architecture. More recent work suggests that perhaps this approach is productive in the short term, but that a longer term solution is to model the physiological substrate that cognition is based upon (Dancy, Ritter, & Berry, 2012; Ritter, Dancy, & Berry, 2011), as well as modeling more complex cognition including multiple types of appraisal and process monitoring.

## Christian Lebiere/Ion Juvina/Alessandro Oltramari

Cognitive architectures such as ACT-R (Anderson & Lebiere, 1998; Anderson et al., 2004) have been quite successful at formalizing and organizing basic cognitive processes in computational frameworks that can accomplish complex tasks. Contrary to common descriptions as purely symbolic or rational, they actually integrate both explicit and implicit cognitive processes, including declarative and procedural knowledge as well as symbolic and subsymbolic levels of representation. The duality of System 1 (automatic) vs. System 2 (controlled) processes is thus an oversimplification of the reality of complex cognition, which integrates basic, intuitive steps of cognition driven by the subsymbolic parameters of the knowledge structures involved into controlled threads of execution capable of accomplishing complex tasks. Despite the success of this purely cognitive approach at modeling a broad range of cognitive tasks, we have found it necessary to contemplate integrating emotional processes into the architectural framework.

The mainstream approach (e.g., Gratch, Marsella et al., 2009; Marinier, Laird, et al., 2009) is concerned with modeling discrete emotions as they arise from appraisal processes that are hardwired in the architecture. Our approach to modeling affective processes is complementary to that approach. We claim that only psychological primitives need to be included in the architecture. Psychological primitives are basic mechanisms that allow us to learn and adapt to the environment. Perceptual experiences, knowledge and skills are not to be included in the architecture. They can be part of specific models and are usually developed through various learning mechanisms. According to this view, discrete emotions are not psychological primitives. They are not biologically given (Barrett, 2006) but instead develop (are learned) from core affect. We conceive of emotion as a perceptual-conceptual experience that is analogous to color perception. People use category knowledge about color to shape the perception of wavelengths of light into the experience of color (Barrett, 2006). Correspondingly, people use category knowledge about emotion to shape the interoception of core affect into the experience of emotion. Core affect is the constant stream of transient alterations in an organism’s neurophysiological

state that represent its immediate relation to the flow of changing events (Russell, 2003). It is typically characterized along two (or three) dimensions: valence and arousal (and approach-avoidance). Changes in core affect can result from physiological (e.g., hunger) and cognitive processes (valuation). Valuation is the process of learning the (expected) value of stimuli encountered in the environment. Very few stimuli have intrinsic value (i.e., they act directly on our nervous system without involving prior learning). Typically, people learn the value of stimuli by associating them with core affect states and external events.

We have developed a simple valuation mechanism that associates a specific value to every representation (chunk). These values are called valuations and can be used to evaluate new stimuli. They are learned via a reinforcement learning mechanism similar to the mechanism of learning the utilities of actions. Thus, the valuation of a chunk is a learned expectation of the likelihood that the chunk would be relevant to the current situation. The relevance indicated by valuation is additive to that indicated by activation. The sign and magnitude of valuation can be used as constraints on retrieval. Valuations are computed based on the rewards that the model receives during its execution and they change as the model is executed.

We claim that activation and valuation (together with learning) are the necessary and sufficient architectural building blocks of cognitive and affective processing. We are using these mechanisms to develop specific models in which cognition and affect interact to produce human-like goal-directed adaptive behavior. For example, in a variant of the game Prisoner's Dilemma, we showed that a cognitive model was more effective than the human participants. Specifically, it learned that cooperation was more beneficial in the long term, and it did not react to occasional unreciprocated attempts to cooperate (Juvina, Lebiere, et al., 2011). However, human participants showed signs of emotional reactivity. Particularly, they were more likely to immediately react by defecting after unreciprocated cooperation, ignoring the potential long-term benefits of sustained cooperation. This behavior has been observed in other studies with similar tasks and associated with a specific pattern of neural activity (e.g., Rilling, 2008). In order to correct for the mismatch between model and human data, we introduced an emotional bias in the model. The assumption was that such a bias develops in human-human interactions to prevent exploitation of a player by another. We claim that such emotional biases are learned from interaction experience using the architectural mechanism described above.

### **Jonathan Gratch**

As someone that studies and models emotion, I definitely agree there is value in a symposium on "beyond rational" processes, but I will take issue with the perspective that attempts to dichotomize cognition and characterizes traditional/successful cognitive modeling as sequential and deliberative. In general, I have a problem with dual process

explanations which (in my view) tend to overly simplify cognition as either: emotional vs. rational; intuitive vs. deliberative; or "System 1" vs. "System 2. Rather, I will argue that dual-process distinctions are largely an artifact of how we study and formalize cognition. On the one hand, normative frameworks for formalizing cognition (e.g., decision theory, game theory or Bayesian inference) highlight human departures from "rational behavior" that may say more about the limits of our frameworks than the duality of human cognitive processes (e.g., see Gigerenzer, 1991). On the other hand, experimental paradigms that illustrate such dualities present participants with unnatural situations designed to highlight these distinctions. Instead, I see thought arising from a tight coupling and dynamic unfolding of a variety of processes (some more naturally characterized as automatic/parallel and some more naturally characterized as sequential).

I am also not convinced that cognitive models are most naturally seen as simply sequential/deliberate. Even early cognitive architectures such as Soar (Newell, 1990) have this close coupling of "automatic" (e.g., elaborations) and deliberative/sequential (e.g., operators) processes (although we might quibble about if this maps well onto any specific dichotomy), and many "successful" cognitive models (e.g. Thagard's 2002 coherence models); models of perceptual or motor processes) are not naturally viewed as sequential.

Despite these quibbles about dual process models, I fully agree that cognitive science, and especially the cognitive modeling community, have largely ignored modeling problems that involve emotion and motivation with the consequence that, on the one hand, we are sorely lacking when it comes to information processing accounts of emotional processes. On the other hand, cognitive models tend to overlook a whole class of problems and mechanisms that might give a different window on how cognition works outside the emotionally-sheltered laboratory.

### **Richard Young (Discussant)**

Richard Young has a long-standing interest in cognitive modeling, cognitive architectures, and related matters. He will respond to the presentations in the symposium, doing his best to identify common threads and contentious themes, before opening the discussion to the audience.

### **Acknowledgments**

This work of the first author was supported in part by AFOSR/AFRL grant FA9550-10-1-0385, the Center of Excellence in Neuroergonomics, Technology, and Cognition (CENTEC), by the Center for Social Complexity at George Mason University, and by the Office of Naval Research (ONR) under a Multidisciplinary University Research Initiative (MURI) grant no. N00014-08-1-0921. The work presented by Frank Ritter was supported by ONR (N00014-03-1-0248) and the Defense Threat Reduction Agency (DTRA). Christian Lebiere (joint work with Ion Juvina and Alessandro Oltramari) is supported by DTRA grant number: HDTRA1-09-1-0053 to Christian Lebiere. Jonathan Gratch

is supported by AFOSR grant FA9550-09-1-0507 and the National Science Foundation grant IIS-0916858.

## References

- Anderson, J. R. (2007). How Can the Human Mind Occur in the Physical Universe? Oxford: Oxford University Press.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglas, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review* 111(4): 1036-1060.
- Dancy, C. L., Ritter, F. E., & Berry, K. (2012). Towards adding a physiological substrate to ACT-R. In *Proceedings of the 21th Conference on Behavior Representation in Modeling and Simulation*. BRIMS Society: Amelia Island, FL.
- Feldman Barrett, L. (2006) Are Emotions Natural Kinds? *Perspectives on Psychological Science*, 1, 1, 28-58.
- Evans, J.St.B.T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology*, 59, 255-278.
- Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases". *European Review of Social Psychology* 2(1), 83-115.
- Gigerenzer, G. (2007). Gut Feelings: The Intelligence of the Unconscious. New York, NY: Penguin Books.
- Gratch, J. & Marsella, S. (2004) A Domain-independent Framework for Modeling Emotion. *Journal of Cognitive Systems Research*, 5(4), 269-306.
- Gratch, J., Marsella, S., Wang, N., & Stanokovic, B. (2009). Assessing the validity of appraisal-based models of emotion. *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*.
- Gray, W. D. (2007). Composition and control of integrated cognitive systems. In W. D. Gray (Ed.), *Integrated models of cognitive systems*, 3-12. New York, NY: Oxford University Press.
- Irving, W.B. (2006). On Desire: Why we want what we want. Oxford: Oxford University Press.
- James, W. (1892/2001) *Psychology: The Briefer Course*. Dover edition republication of Harper Torchbook edition 1961 abridged of original work first published by Henry Holt and Co. 1892. Toronto, Ontario: Dover edition.
- Jovina, I., Lebiere, C., Martin, J.M., & Gonzalez, C. (2011) Intergroup Prisoner's Dilemma with Intragroup Power Dynamics. *Games*, 2, 21-51.
- Kahneman D. (2003). A perspective on judgment and choice. *American Psychologist*. 58, 697-720.
- Klein, G. (1998). Sources of power: How people make decisions. Cambridge, MA: The MIT Press.
- Kennedy, W.G. & Bugajska, M. (2010). Integrating Fast and Slow Cognitive Processes. In D.D. Salvucci & G. Gunzelmann (Eds.), *Proceedings of the International Conference on Cognitive Modeling (ICCM 2010)* (pp. 121-126). Philadelphia, PA, (4-8 August 2010)
- Laird, J.E. (2008). Extending the Soar Cognitive Architecture. *Proceedings of the Conference on Artificial General Intelligence*. Memphis, TN: ISO Press.
- Laird, J.E. (2012 in press). *The Soar Cognitive Architecture*. Cambridge, MA: The MIT Press.
- LeDoux, J.E. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Simon & Schuster.
- Marinier, R.P., Laird, J.E., & Lewis, R.L. (2009) A computational unification of cognitive behavior and emotion. *Cognitive Systems Research*, 10, 48-69.
- Minsky, M. (2006). *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York, NY: Simon & Schuster.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Paivio, A. (1971) *Imagery and verbal processes*. New York, NY: Hold, Rinehart, & Winston.
- Rilling, J.K. (2008) Neuroscientific Approaches and Applications within Anthropology. *Yearbook of Physical Anthropology*, 51, 2-32.
- Ritter, F. E., Dancy, C., & Berry, K. (2011). The case for including physiology in cognitive modeling. In *Biologically Inspired Cognitive Architectures*, 303. Amsterdam: IOS Press.
- Russell, J.A. (2003) Core Affect and the Psychological Construction of Emotion. *Psychological Review*, 110, 1, 145-172.
- Simon, H.A. (1982). *Reason in Human Affairs*. Stanford University Press.
- Slovan, S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.
- Thagard, P. (2002). *Coherence in Thought and Action (Life and Mind: Philosophical Issues in Biology and Psychology)*. Cambridge, MA: The MIT Press.
- Thagard, P. (2006). *Hot Thought: Mechanisms and Applications of Emotional Cognition*. Cambridge, MA: The MIT Press.
- Wallach, D. & Lebiere, C. (2003). Implicit and explicit learning in a unified architecture of cognition. In L. Jimenez (Ed.) *Advances in Consciousness Research*, 215-250. Amsterdam: John Benjamins Publishing Company.