Scaling Up High-Fidelity Cognitive Modeling to Real-World Applications

Christian J. Lebiere, Andrea Stocco, David Reitter, Ion Juvina
Psychology Department, Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh PA 15213
USA

cl@cmu.edu; stocco@cmu.edu; reitter@cmu.edu; juvina@cmu.edu

ABSTRACT

The approach presented in this paper addresses the question of the proper scientific basis for Human Factors modeling and proposes an architectural framework for integrating federated models and simulations. It is intended to be applicable to a broad range of scenarios across domains. Indeed, broad applicability and integration of modeling and simulation techniques are essential to their effectiveness and validation. Our approach is grounded in the concept of unified theories of cognition, implemented computationally as cognitive architectures. However, cognitive architectures also need to be constrained by our knowledge of neural processes in order to properly account for all cognitive, perceptual and motor factors. Despite that attention to the small-scale basis of cognition, cognitive modeling can scale up to social situations and large-scale network settings through a process of abstraction and integration. Key to that process is the availability of easily accessible resources in the form of existing cognitive models, implemented tasks, simulation environments adhering to a common standard, and human performance data to constrain and validate models. Investment in that infrastructure is essential to ensure growth and scalability in the application of cognitive models and their proper integration in military simulations.

1.0 INTRODUCTION

Human behavior models have proliferated recently, driven by the need to quantitatively predict the effectiveness of systems and courses of actions in situations where the human element is determinant. However, while the multiplication of models has led to specific successes, it has not generally resulted in the kind of incremental progress in their accuracy and effectiveness that had been expected.

We argue that the reason for the limited success reached to date originates from the focus on maximizing the fit of specific models to narrow data sets, instead of incrementally broadening existing accounts and integrating results into increasingly pervasive and accurate models of human behavior. A major issue is the need to cope with the multiple levels of analysis of human behavior, which span highly detailed models of individual behavior at the sub-second time scale to large-scale abstract models of behavior at the group level. Unfortunately, those levels of description have typically little interaction between them, and lack the reductionist approach that characterizes the hard sciences, where each level strongly constraints the more abstract levels built upon it. It is therefore essential to ground the larger-scale models of human behavior in the individual acts of cognition from which they emerge.

Pursuing an integrated approach raises a number of challenges. First, we argue that cognitive models need to be grounded and constrained by our knowledge of neural processes and their interaction in order to be able to provide the true scientific basis for the entire edifice that rests upon them. However, that emphasis on the small-scale only exacerbates the problem that high-fidelity cognitive models already strain our computational

Report Documentation Page			*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.			
1. REPORT DATE **OCT 2010**	2. REPORT TYPE **N/A**		3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Scaling Up High-Fidelity Cognitive Modeling to Real-World Applications**			5a. CONTRACT NUMBER
			5b. GRANT NUMBER
			5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)			5d. PROJECT NUMBER
			5e. TASK NUMBER
			5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Psychology Department, Carnegie Mellon University 5000 Forbes Avenue Pittsburgh PA 15213 USA**			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited**			
13. SUPPLEMENTARY NOTES **See also ADA564696. Human Modelling for Military Application (Applications militaires de la modelisation humaine). RTO-MP-HFM-202**			

14. ABSTRACT

The approach presented in this paper addresses the question of the proper scientific basis for Human Factors modeling and proposes an architectural framework for integrating federated models and simulations. It is intended to be applicable to a broad range of scenarios across domains. Indeed, broad applicability and integration of modeling and simulation techniques are essential to their effectiveness and validation. Our approach is grounded in the concept of unified theories of cognition, implemented computationally as cognitive architectures. However, cognitive architectures also need to be constrained by our knowledge of neural processes in order to properly account for all cognitive, perceptual and motor factors. Despite that attention to the small-scale basis of cognition, cognitive modeling can scale up to social situations and large-scale network settings through a process of abstraction and integration. Key to that process is the availability of easily accessible resources in the form of existing cognitive models, implemented tasks, simulation environments adhering to a common standard, and human performance data to constrain and validate models. Investment in that infrastructure is essential to ensure growth and scalability in the application of cognitive models and their proper integration in military simulations.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**20**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

resources. We describe a methodology that scales up cognitive models to increase their applicability to large-scale simulations while preserving the key aspects of their fidelity. Scientifically, we need to extend the top-down cognitive modeling approach that has overwhelmingly been developed and applied in single-agent tasks to account for emergent effects in cultural and social interactions. We describe how those same cognitive mechanisms validated in laboratory experiments can explain key aspects of social interactions.

A third obstacle originates in the fragmentation of the practice of cognitive modeling. The standard practice remains to develop single-task models that do not generalize to other tasks and are not integrated with each other. We advocate a pragmatic solution that has been successfully applied in numerous fields from physics to genetics where it has led to rapid incremental progress. The solution consists in creating a repository of cognitive architectures, computational models, implemented task environments and human performance data sets that can support reuse and sharing of models. In turn, this can lead to the integration of increasingly complex models of human behavior, validated over a broad set of empirical results across multiple fields. Such a solution will require fundamental advances both in combining cognitive models and providing a general framework for interacting with a full range of simulation environments. But it promises to deliver the benefits oh high-fidelity cognitive models to a new range of applications.

2.0 COGNITIVE ARCHITECTURES

Theoretical unification is the goal of all sciences. In reaction to the Newell's claim that the study of human cognition was growing increasingly fragmented and to his suggestion that a unified computational framework was needed to bring the field together (Newell, 1973a; 1973b), a number of unified theories of cognition, implemented computationally as cognitive architectures, have been proposed (e.g., Anderson 1983; Meyer & Kieras, 1997; Newell, 1990; Sun, 2007). We focus here on one particular cognitive architecture, ACT-R (Anderson & Lebiere, 1998; Anderson, 2007), noting a gradual convergence in the field toward an increasingly accepted set of mechanism and organizational principles.

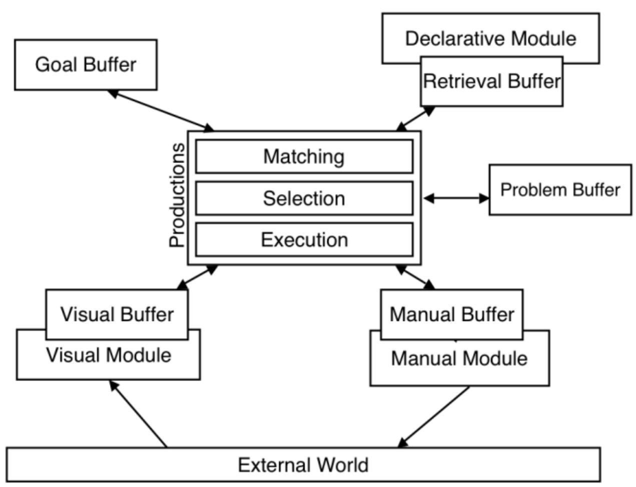

Figure 1: Overview of the ACT-R Cognitive Architecture.

The primary architectural commitments of ACT-R are two-fold. At the organizational level (see Figure 1), the architecture is composed of a set of modules, including perceptual (visual), motor (manual), and declarative memory modules (as well as self-standing goal and imaginal (problem) buffers), coordinated by the procedural (production rules) module through limited-capacity buffers. Each processing step within a module is massively parallel (e.g., all production rules in the procedural module are matched at once, as are chunks of information in the declarative module) while communication between modules is serial and asynchronous (e.g., only one request for information retrieval can be sent to declarative memory at a time, and a single chunk will be returned through the retrieval buffer when the retrieval is completed). Activity in the modules is correlated with fMRI BOLD response in specific brain regions, bringing to bear neuroscience constraints on architecture organization (Anderson, 2007).

The second level of architectural commitments concerns the representations and processes taking place in each module, and in particular the declarative memory and the procedural modules. ACT-R's approach is a hybrid combination of a simple, constrained symbolic representation (chunks in declarative memory, production rules in the procedural module) together with subsymbolic selection mechanisms that adapt to the statistical structure of the task and its environment. The former underlies our ability to perform almost any task and quickly learn novel combinations of knowledge while the latter captures the soft, adaptive nature of human performance including both its abilities (e.g., generalization) and its limitations (e.g., forgetting). A tight integration of those two very different types of abilities is necessary to account for the full range of human cognition (Anderson & Lebiere, 2003). A full list of the tasks and domains modeled using the architecture is available on the ACT-R web site (http://act-r.psy.cmu.edu). Topics range from perception and attention to problem-solving and decision-making while tasks range from basic psychology experiments in memory to complex simulations such as air-traffic control and driving a car in traffic.

A general feature of scientific theories is that they explain phenomena in terms of simpler, underlying mechanisms. ACT-R not only provides a cognitive explanation of the wide range of behavioral phenomena described above at level of basic cognitive representations and processes, it has also successfully incorporated theoretical postulates that offer explanations of its cognitive mechanisms at two finer levels of abstraction. Specifically, a sub-symbolic level of operation (Lebiere & Anderson, 1993; Lebiere, Anderson & Reder, 1994) has been added that expands its explanatory power to phenomena of the type that models emphasizing parallel distributed processing (Rumelhart & McClelland, 1986) have been especially successful in explaining. Second, recent efforts have further extended the architecture's scope by linking key cognitive functions to brain mechanisms (Anderson, 2007). Significant progress has been made in accounting for performance in temporally extended tasks not only at the level of cognitive processes but also in predicting the brain areas activated by these process as they are executed (Anderson, 2009). Such layered explanations are like explanations of biological processes at molecular and atomic levels. In short, ACT-R organizes a wide range of phenomena, generates a wealth of testable predictions, and demonstrates considerable explanatory power.

3.0 THE NEURAL LEVEL

3.1 Integration Across Levels

As suggested above, in addition to the integration of models "horizontally", i.e. across tasks and domains, we also advocate for an integration of models "vertically", i.e. across different levels of description. Unfortunately, integration among levels of description has been traditionally as neglected as the horizontal integration across domains, leading to frameworks and theories that grow in isolation and whose implications are often not well understood at other levels of analysis.

One might argue for a separation of levels of analysis, and maintain that understanding the biological basis of a process does not give much more information. This is, however, not the case. In the study of human behavior, any phenomenon ultimately has a neurological basis, and lower-level models help constrain higher-level descriptions (Newell, 1980). In several occasions an analysis in terms of basic units of computations in the brain (that is, neuronal ensembles) has provided notable insights into our understanding of a cognitive, behavioral, or even social process.

Thus, the first benefit of integration across levels is that it brings additional constraints. At any level of analysis, different descriptions of the same phenomenon are possible, and many of them might be equally compatible with the empirical data. Breaking down the description into its basic components can obviously help discriminate between the alternatives. Sometimes, in fact, it can even suggest new and better descriptions that have not been thought of previously. And sometimes an analysis at a lower level can get rid of concepts that just do not reflect the underlying reality, or eliminate false dichotomies that hide a continuum.

3.2 The Conditional Routing Model

Our work one the conditional routing model (Stocco, Lebiere, & Anderson, 2010) provides a good example of these implications. The conditional routing model was initially developed to bridge the gap between production systems (a common formalism for cognitive processes) and more biological descriptions of individual brain circuits, such as neural networks. Production systems work by executing rules in the form of IF-THEN constructs, and have long become a standard tool for modeling behaviors (Newell, 1973), from complex task to laboratory experiments in cognitive psychology. However, they seem to have very little in common with the graded and parallel computations occurring in the brain.

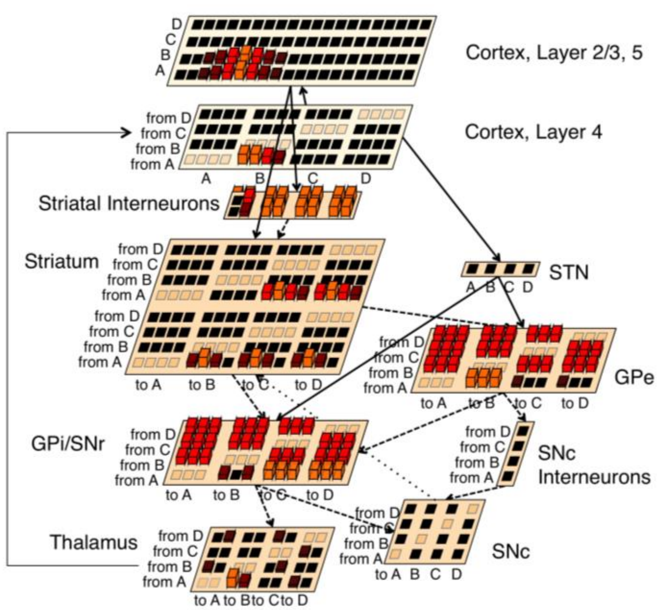

Figure 2: Overview of the Conditional Routing Model.

Work on the routing model, however, brought to light unexpected similarities between the operations of a production system and the cycles of activity of the basal ganglia (see Figure 2 for details). The applications of rules had its counterpart in the opening or closing of channels that connect the basal ganglia with the prefrontal cortex. Perhaps the most interesting result concerned the computational concept of variable. Part of the strength of production rules comes from their use of variables, which make them applicable to a general class of conditions and also permits manipulation of values. The routing model accounts for computations that correspond to variables in the human brain.

Importantly, the conditional routing model had implications that scaled up to higher levels of analysis. It provided a rationale for the use of production systems in modeling behavior. It also provided novel constraints to production system-based models. For instance, different cognitive architectures have disagreed on whether the execution of production rule constitutes a central bottleneck (as in ACT-R: Anderson, Taatgen, & Byrne, 2005) or an unlimited resource that can be accessed in parallel (as in EPIC: Meyer & Kieras, 1997). The routing model puts the problem in a novel perspective, and shows how the ultimate constraint is a matter of channel capacity, i.e. not how many rules can be executed, but how many signals and how much information can be transferred at the same time.

At an even higher level of analysis, the conditional routing model provided the basis for understanding one of the central questions in neuroscience. Simulations showed that the basal ganglia circuit can be controlled by regions that reflect the topology of brain connections. Since it could control the basal ganglia circuit, such a region would be in an ideal position to control the channels that transfer signals between brain regions, and, ultimately, control how behavior would be executed. This region would in fact hold representations of cognitive actions to be performed, such a plans or verbal instructions.

In fact, we run a neuroimaging experiment and found evidence for this processes. We identified a candidate region that is recruited when novel instructions are given, and we were able to show that the basal ganglia work as an instruction interpreter. Although work is still in progress, this work promises important contributions in all the fields where behavior is shaped by instructions, from education to military missions.

4.0 THE SOCIAL LEVEL

4.1 IPD2: A Game Paradigm for Studying the Intragroup Power Dynamics

The overwhelming majority of experimental paradigms in cognitive psychology involve a single subject engaged in a specific task using a computer or similar apparatus where the course of the task, e.g., the situations that are encountered, has been determined in advance by the experimenter. Thus theories of cognition have arisen to account principally for those situations, and have not usually concerned themselves with the issues that are involved when interacting in social situations, such as cultural differences or emotional reactions to other people’s actions, and how they might influence our behavior.

Game theory offers simple paradigms through which our decisions under conditions in which our interest might conflict with those of others can be studied systematically (e.g., Rapoport, Guyer & Gordon, 1976). Lebiere, Wallach & West (2000) showed that the same general cognitive mechanisms used in single-agent tasks, such as memory and decision-making, could account for a broad range of results. Yet it is undeniable that emotions play a role in influencing human decisions in social situations.

However successful game theory has been in capturing some of the dynamics involved in inter-personal conflicts, its typical paradigms often lack the structure present in many of the challenges faced in political, economic and military situations that arise in the real world. Thus, we have created a generalization of the typical 2x2 paradigm of game theory called IPD2, for Intergroup Prisoner's Dilemma with Intragroup Power Dynamics. IPD2 is intended to improve the ecological validity of the existing Iterated Prisoner's Dilemma (IPD) paradigms. It reflects the situation in which players strive to achieve not only money/payoff (as in the classical economic theory) but when they also have a more complex set of motives. In this version of the game we add the power motive. We define power as the ability to have a say in (contribute to) what happens to one's group. Evidently, power does not equate with payoff. Free riders get payoff without having power. Correspondingly, having power does not guarantee making profitable decisions. However, having a certain level of power can be a prerequisite for achieving significant amounts of payoff, and in return, obtaining payoff can be a means to retain or achieve greater power.

As a consequence of the complex interplay between power and payoff, a realistic game like IPD2 departs from the classical behavioral game theory paradigm in which a payoff matrix will be either explicitly presented to the participants in advance or easily learned from the experience of game playing. The players might not be able to figure out the full range of game contexts and outcomes. They might have to satisfice and use heuristics, which is what humans in the real world do (Gigerenzer & Todd, 1999).

In IPD^2, there are two teams of two players playing an inter-group Prisoner's Dilemma game. Other payoffs matrices would also be possible, but we decided to use the Prisoner's Dilemma because of its widespread applicability to adversarial inter-group situations. Within a team, both players choose between cooperate (C) and defect (D) options, but only one player's choice counts as the choice of the team. What determines whether a player's choice counts as the team's choice is that player's "power", specifically whether its power is greater than the power of the other team member. The player whose choice counts is said to be in the majority while the other is in opposition, by analogy to electoral power, though this concept could account for other forms of power as well, such as demographic or military advantage. A player's power is a quantity assigned at the start of the game and increased or decreased after each round of the game depending on the player's choice, the teammate's choice, and the opponent team's choice. The sum of power within a team remains constant throughout the game at 1.0. All players start the game with the same amount of power, 0.5. A random value is added or subtracted from each player's power level at each round. This random noise has the functions of breaking ties (only one player can be in power at any given time) and preventing quick strategizing and settling into fixed behaviors.

After a round of simultaneous decisions, if the two members of a team made the same decision (both played C or both played D), their powers do not change other than for the random factor. However, if they made different decisions, their powers change in a way that depends on the outcome of the inter-team game (PD in this case). A fraction of the payoff received from the payoff matrix is added to the power of the player in the majority and the same amount is subtracted from the power of the player in opposition:

$$power(t) = power(t-1) + \frac{team\_payoff\,(t)}{100} \qquad \text{if player is in the majority}$$

$$power(t) = power(t-1) - \frac{team\_payoff\,(t)}{100} \qquad \text{if player is in opposition}$$

where *power(t)* is the current power, *power(t-1)* is the power from the previous round, and *team_payoff (t)* is the current team payoff from the inter-team game. The scaling factor of 100 is used to convert from a payoff scale typical of the PD game (e.g., -10, -1, 1 and 10) to a power scale expressed as a fraction of one unit. Note that the values in the payoff matrix can be positive or negative. Thus, if the team receives a positive payoff the power of the player in the majority increases whereas the power of the player in opposition decreases by the same amount. If the team receives a negative payoff, the power of the player in the majority decreases whereas the power of the player in opposition increases.

A player's power does not only allow a player's decision to count as its team's decision. It is also factored in each player's payoff. The payoff that the team receives is shared between the two teammates in direct proportion with their power. Thus, for both the player in power and the player in opposition, individual payoff increments as:

$$payoff(t) = payoff(t-1) + power(t) * team\_payoff(t)$$

where *payoff(t)* is the current individual payoff and *payoff(t-1)* is the previous individual payoff.

Again, since the team's payoff can be positive or negative, differences in individual payoff between teammates can be directly or inversely proportional to differences in power. For example, if the team gets a negative payoff, then the individual payoff of the majority player is decremented by a larger amount than the individual payoff of the opposition player.

4.2 Empirical Study and Cognitive Model

We conducted an empirical study aimed at exploring the potential of the game to represent realistic human behavior. We recruited 68 participants and paired them with computer strategies of various complexities. The following are descriptions of the strategies that were used in this study:

- *Always-cooperate* and *always-defect* were the two maximally predictable strategies.
- *Tit-for-tat*, a classic PD strategy, starts by cooperating and then repeats the last choice of the opposing team.
- *Seek-power* plays differently depending on its power status. When the player is in power, it plays tit-for-tat. When in opposition, it plays C or D depending on the outcome of the intergroup Prisoner's Dilemma game at the previous iteration. If this outcome is symmetric (CC or DD), seek-power plays C, if it is asymmetric (CD or DC), seek-power plays D. Either way, the intention is to gain power if the previous outcome repeats itself.
- *Exploit* plays "win-stay, lose-switch", another classic game theory strategy, when in power and seek-power when in opposition.

Some of these strategies are very simple and easily predictable, others are more complex and harder to predict. The intention was to give humans a large enough set of strategies to play against so as to manifest realistic behavior. The human participants were matched with each strategy both as teammates and as players on opposing teams. As a result, ten game types were constructed, representing all possible combinations of 3 of the 5 strategies. Each human participant played a practice game followed by the ten game types. A Latin square design was used to counterbalance the order of game types. Each game was played for 50 rounds. Participants were instructed to try to maximize their payoff.

We developed a cognitive model inspired by the Instance Based Learning theory (Gonzalez, Lerch, & Lebiere, 2003) implemented in the ACT-R cognitive architecture. All the parameters of the ACT-R architecture were kept at their default values. For each round of the game, the model tries to remember if it has encountered the same situation (context) in the past and what was its action in that situation. A context is characterized by the choices of all the players and the group choices in the previous round. If such a context-action pair is retrieved from memory, the model takes the retrieved action. If a context-action pair is not retrieved, the model repeats its previous action and saves it in memory together with the current context. Once an action is taken (a choice is made) the model receives positive or negative payoff. When the payoff is negative, the model saves the current context together with the alternative action. Thus, after receiving negative payoff, the model has two instances with the same context and opposite actions. Retrieving and creating context-action pairs makes them more active and more available for retrieval when needed, as constrained by the ACT-R theory of memory (Lebiere, Wallach & West, 2000). A time-based decay process causes the existing memories to be forgotten if they are not frequently used.

Humans are able to learn the game and perform adaptively, that is, they manage to get positive payoff by the end of the game, averaged over all 10 game configurations (Figure 3). However, the learning curve is different than the typical learning curve found in single-agent non-interactive tasks. The typical learning curve in such tasks starts with a steep increase until it reaches a plateau. The learning curve found here starts with a quick decline and plateau in the first half of the game and ends with a steep increase toward the second half of the game. The model captures these dynamics as a function of its reliance on memory. At the beginning of the game, the model does not have enough instances in its memory to make correct choices, and thus plays quite predictably. As the game progresses, more instances are accumulated, the more effective ones become more active, and the ineffective ones are forgotten.

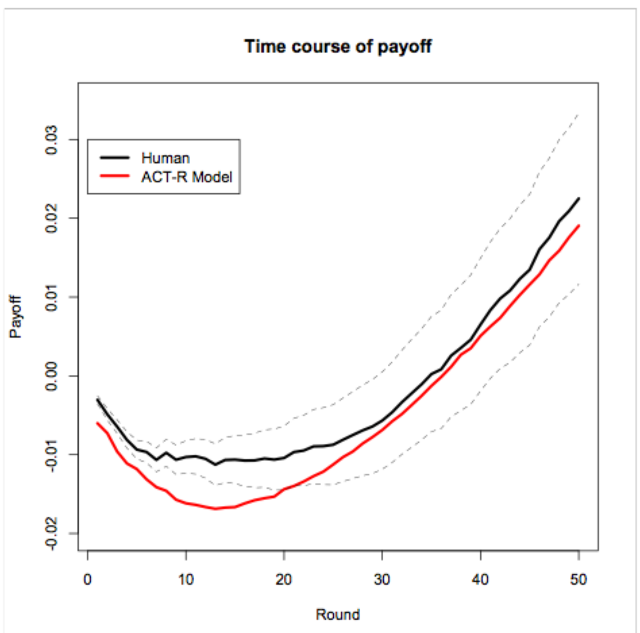

Figure 3: Time Course of Payoff.

Human participants manage to increase their power over time and the model shows the same trend (Figure 4). The participants were not explicitly told to try to increase their power and the model is not sensitive to changes in power. Thus, the increase in power was probably a byproduct of the focus on maximizing payoff.

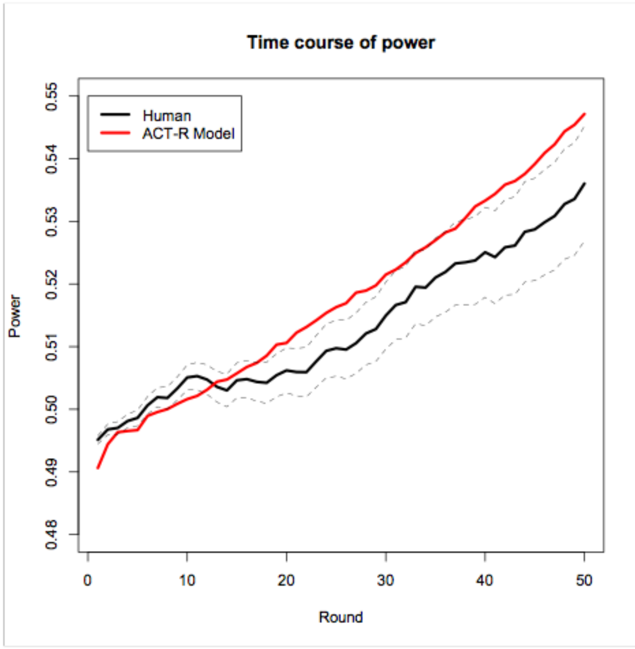

Figure 4: Time Course of Power.

Regarding specific choices made, humans slightly increase their cooperation throughout the game. The model shows the same trend but a higher level of cooperation overall (Figure 5).

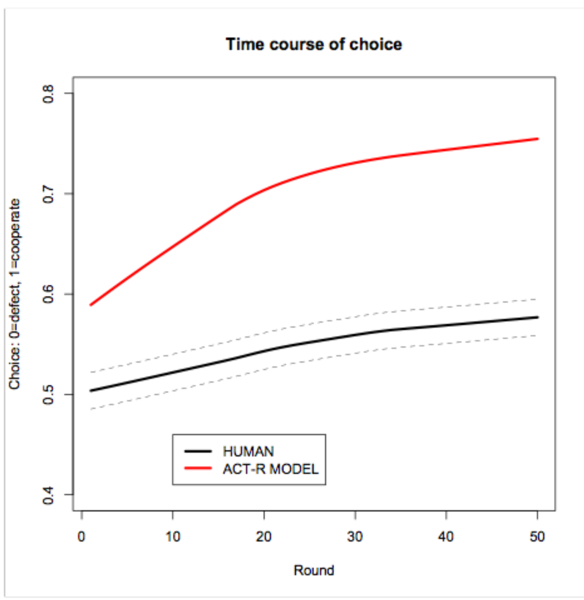

Figure 5: Time Course of Choice.

We have shown that intragroup power dynamics in conditions of intergroup conflict/cooperation can be explained by simple cognitive mechanisms. However, the model predicts higher levels of cooperation than observed in the human data. Humans seem to be more risk-seeking or greedy than the model. One limitation of the model is that it is not sensitive to the magnitude of the payoff but only to its valence (positive or negative). This limitation could be responsible for the failure of the model to account for the amount of cooperation in the human data. We are currently looking into ways to make the model sensitive to the magnitude of the payoff. One possibility is the use of blended memory retrieval to generate quantitative expectations, as has been used successfully in a model of probabilistic choice (Erev et al, 2010).

5.0 THE NETWORK LEVEL

5.1 Scaling Up Cognitive Modeling

Cognitive models have explained a great deal of behavioral and neurophysiological data. On the road to understanding the mind, cognitive architectures have specified a core set of representations and mechanisms common to a variety of models in order to separate general functional components and their abilities from domain-specific instantiations, such as knowledge and strategies. However, the tasks that classical cognitive models have taken on are mainly those that can be defined in a controlled environment. Process models of laboratory behavior are often overly specific and needlessly complex, while alternative models would yield similar fits. The model eco-system has diversified rather than converged, with specific rule sets developed for each given task and very seldom reused or generalized for other tasks. This leads to overfitting and lack of robustness. To robustly explain and predict behavior in complex real-life situations, model complexity has to increase further. Inevitably, humans execute much more complex tasks as well, drawing from a variety of knowledge and skills and contextualizing their observations and thoughts in light of both long-term experience and recently acquired knowledge.

The greater complexity of tasks may have a welcome effect on cognitive architectures. Current general architectures such as ACT-R (Anderson, 2007) or SOAR (Laird, 1987) are not as restrictive as human cognition is. ACT-R, has, during the transition from versions 2 (Anderson, 1993) to 4 (Anderson & Lebiere, 1998) and 6 (Anderson et al, 2004), become more and more restrictive: large, very complex rules made way for smaller, granular ones that could perform less functionality each. Similarly, complex representation structures of unbounded complexity gave way to smaller, more limited ones that have to assemble in hierarchies to hold the same knowledge. Other mechanisms such as the goal stack have disappeared entirely in favor of reliance on more error-prone long-term memory.

Still, implausible assumptions remain, such as in the ability of holding contextual information in working memory without decay or interference, or to precisely specify the type and detailed structure of memory items. For instance, one could implement a model that predicts excellent human performance at the most intricate N-back working memory task, failing to explain the dismal human performance scalability at this task (Kirchner, 1958). These simplifications, or lack of constraints, have often been made to facilitate the task of the modeler and maximize their low-level control over models of simple laboratory tasks. However, in the long run this level of control is not scalable to more open-ended, unpredictable environments. Moreover, it tends to prevent the model from generalizing easily to related tasks, and makes it hard to combine models that might have different parameters or control regimen.

As task complexity increases, a careful analysis of the components of the model is necessary. Every rule, every data structure, and every knowledge access process can be seen as a claim that needs to be proved empirically. For anything but the simplest cognitive models, many of the procedures and data structures they define are often not evaluated: the specifics of many of the components of the model may be irrelevant to the claims of the model. The solution to this problem is under-specification. In what we call the *Accountable Modeling* paradigm, we suggest to apply Occam's razor and specify only what is meant to be directly or indirectly evaluated.

As a consequence, we arrive at models that can be more complex yet faster and easier to prototype, while still using the same core representations and mechanisms of the architecture. Until all portions of the model are fully specified, such models may fall short of Newellian complete process models. Yet, they honestly separate claim from conjecture and provide the same level of comparison to human data.

5.2 Accountable Modeling

Recent work has been undertaken to investigate the use of ACT-R to study the interaction of two, eight, or even thousands of cognitive agents. Scalability in this domain would make cognitive models applicable to new domains such as network science, for which a precise computational representation of human cognitive processes has been desirable but as to now unavailable. The modeling methodology in this paper follows accountable modeling within the ACT-R theory. The ACT-R framework defines a component-based architecture, in which specialized modules work largely in parallel to contribute to thought processes. In recent computational implementations, it requires end-to-end models, describing thought processes through a set of production rules controlling the interaction of cognitive (e.g., long-term memory), perceptual and motor components.

Working within the ACT-R theory, we designed a new toolbox instantiation of the theory called ACT-UP. ACT-UP reflects ACT-R, but lets the modeler specify algorithms much like a programmer would. Functionality is compartmentalized in reusable functions (taking arguments and returning a value) and data is stored and retrieved as in ACT-R in chunks in declarative memory. Thus, ACT-UP models take a higher-level perspective of the cognitive processes that must take place in order to execute tasks. But ACT-UP also makes more fine-grained cognitive functions available. Such micro-functions allow models to go beyond what is available to ACT-R models. We intend to address several goals with ACT-UP. *Accountability* suggests under-specifying model components that are neither motivated by data or theory nor subject to empirical evaluation. *Rapid prototyping* allows modelers to quickly build and modify most parts of the model, even computationally complex ones, while focusing on learning and other cognitive effects predicted by ACT-R's theoretical assumptions. Crucially, it produces models that are reconfigurable so that systematic parameter search can be used to explore the space of possible models. *Reusability* results from clear input and output data structures, turning models into functions that can be reused in other contexts: the convergence of models and cognitive frameworks is a long-term goal. *Scalability* allows models to run longer, apply to more complex tasks, and simulate agents in the context of larger multi-agent systems. *Cognitive validity* is addressed by replicating existing modeling results, which has resulted in equivalent performance for simpler and computationally more efficient models (Reitter & Lebiere, 2010). A number of novel cognitive models have been implemented that address multi-agent tasks (Reitter & Lebiere, in press), game-theoretic tasks (Reitter, Juvina, Stocco, & Lebiere, 2010) and predictive modeling of control tasks (Reitter 2010).

5.3 Language Convergence

A language, even if shared among the members of a community, is hardly static. It is constantly evolving and adapting to the needs of its speakers. Adaptivity in natural language has been found at various linguistic levels. Models of the horizontal transmission of cultural information within generations show on a much larger scale how beliefs or communicative standards spread within a single generation of humans. The individual agents that are effecting the language change depend on their cognitive abilities such as memory retrieval and language processing to control and accept novel communication standards. Do the local, cognitive constraints at the individual level interact with the structure of large-scale networks? Both social structure and individual cognitive systems have evolved over a long period of time, leading to the hypothesis that certain network structures are more suitable than others to convergence, given the specific human cognitive apparatus.

Network structure, on a small scale, does influence the evolving patterns of communication. The dichotomy between individual and community-based learning motivated experiments by Garrod, Fay, Lee, Oberlander & Macleod (2007) and Fay, Garrod, Roberts & Swoboda (2010) where participants played the Pictionary game. In each trial of this naming game, each participant is paired up with another participant. One of them is then asked to make a drawing to convey a given concept out of a small set of known concepts; the other one is asked to select the concept from that list without engaging in verbal communication. Over time, participants develop common standards codifying those concepts: they develop a system of meaning-symbol pairs, or signs. We take this system as the lexical core of the shared language. The convergence rate and the actual language developed differed as a function of the structure of the small participant communities: the same pairs of participants engaged in the activity repeatedly (Isolated Pairs condition), or different pairs of participants matched up over time (Communities condition). Fay and Garrod's Pictionary experiments served as the empirical basis for a cognitive process model we developed in ACT-UP.

The model explains the convergence as a result of basic learning and memory retrieval processes, which have been well understood and made available for simulation in ACT-R. Thus, properties of human memory and of the agent's learning strategies dictate how quickly they adopt signs or establish new signs: processes such as learning, forgetting and noise together with their fundamental parameters that are within well-established ranges by current modeling practice provide strong constraints on the behavior of each agent and in turn on the evolution of their communication within the network. This approach acknowledges that cultural evolution is constrained by individual learning; each agent learns according to its cognitive faculty (Christiansen & Chater, 2007). With non-cognitive models, language change has been simulated on a larger scale as well, but without the benefit of those constraints (Kirby & Hurford 2002).

The results (Figure 6) show how the simulated agents improve their communication accuracy (proportion of correctly identified signs) over time, and how they suffer from changing communication partners. The convergence as well as the relative success of isolated pairs reflects well Fay&Garrod's empirical data.

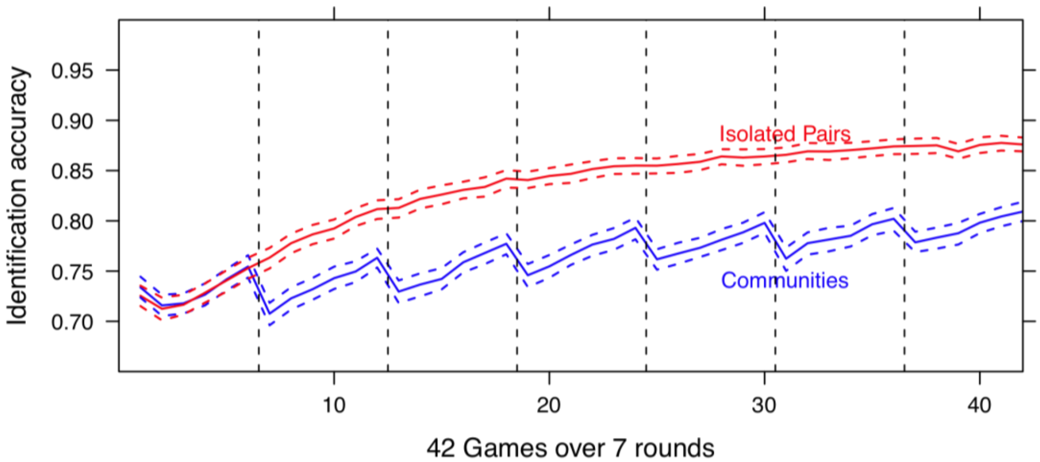

Figure 6: Convergence in Isolated Pairs and Communities Conditions.

Differences in naturally occurring social networks are hardly as extreme as in Fay's experiments. Some agents will be connected to a large number of other ones, while many agents will have just a few connections each. Concretely, the number of interaction partners of a randomly chosen community member is not normally distributed and centered around a mean. It shows a (Zipfian) power law distribution, with a number of hubs attracting many network neighbors, and a long tail of subjects interacting with just a few nodes each. Social networks are small world networks: the average distance between any two nodes in the networks is low, since many of them are connected to hubs, and the clustering coefficient is high (Watts & Strogatz 1998). Non-organically connected communication and command networks follow other normals - tree graphs for instance. However, natural communication standards develop in networks that have very specific properties that can be observed in most organically developed networks.

Because natural language as well as other communication forms are constrained by cognitive function and evolved through a social process, we examined whether human memory may be uniquely adapted to the social structures prevalent in groups, specifically small-world networks. Several community structures were examined (grids, trees, random graphs and small-world networks). We found that convergence is relatively stable across the four network types. Analyzing the differences between the networks, we find that the average degree, which was controlled for grids, random networks and small worlds, was substantially lower for trees (d=1.9) than for the other networks (d=5), due to the large number of leaves with degree 1. This (or the correlated algebraic connectivity of the network) may prove to be a deciding correlate with cross-network convergence.

In a first simulation, we used relatively small networks (85 nodes). Figure 7 shows the learning curve for agent pairs in the four networks. Agents in all networks converge. Confidence intervals obtained via bootstrapping indicated no apparent differences at any specific iteration.

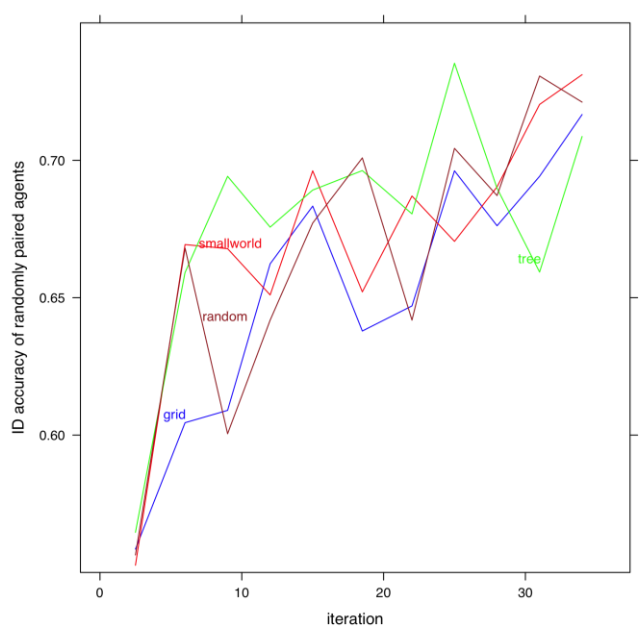

Figure 7: Performance as a Function of Network Structure.

A linear model was fitted estimating the effects of network type overall (as a baseline) for each of the four types. It also fitted interactions of iteration (1–35) with the network types, which indicate significant learning effects as follows. For each network type, we found a significant learning effect (effect of Round) (β 0.002, $p < 0.001$). However, individual communication success does not indicate convergence across the network. How would randomly matched pairs perform after training within the established network? Scaling up the network to 512 nodes (d=6) allowed us to examine this question in a second simulation. We measured communication accuracy between pairs of randomly chosen agents after each round. For three network types, Grid, Small World and Random we found significant interactions with round, i.e. significant convergence, (all $\beta > 0.016$, $z > 2.1$, $p < 0.05$). For the network type Tree we found no significant interaction ($\beta = 0.012$, $z = 1.55$, $p = 0.12$). To test the initial hypothesis, we re-coded the conditions with a SmallWorld factor, contrasting the small world networks with all other conditions. We found an effect of Round ($\beta = 0.017$, $z = 3.66$, $p < 0.001$), indicating convergence, but no interaction with SmallWorld ($\beta = -0.00027$, $z = -0.03$, $p = 0.98$). Our ability to quickly prototype cognitive models and scale them up to network simulations involving large numbers of agents played a crucial role in establishing this link between individual cognitive properties and overall network performance, and more generally in applying high-fidelity cognitive modeling techniques to the field of network science.

6.0 BRINGING MODELS AND SIMULATIONS TOGETHER

6.1 Integrated Repository of Models, Simulations, Tasks and Data

The increase in computational power and software sophistication in recent decades has enabled the growth of increasingly complex models of cognitive and brain functions. Increasingly complex models per se, however, are not a guarantee of progress. Models can be developed at different levels of abstraction, making their relationship to each other unclear until clear links between levels are established. Modeling paradigms also tend to specialize to particular classes of tasks for which they are well suited, assuming but seldom establishing their applicability and relevance to other types of tasks. Sets of mechanisms and representations are often posited and bundled together, making credit assignment of successes to individual components difficult to perform, and generalization difficult. While models accumulate, true cumulative progress remains elusive.

The objective of an integrated repository is to identify the necessary means to achieve greater rates of convergence and incremental progress in cognitive modeling through the use of a shared repository of computational cognitive models, experimental tasks, and performance data. This repository would serve multiple complementary purposes.

First, an integrated repository would facilitate direct comparison of different models. The development and study of cognitive models has been ongoing for decades, yet it is difficult to see how different models map onto each other, what features or components are missing, and what progress has been made. Different modeling communities speak different languages and largely ignore each other. Detailed models and data are seldom available in a comparable format, making direct model comparisons partial at best and tendentious at worst. Therefore, a common language for the description of modeling paradigms is needed to develop an understanding of how they relate to each other. An integrated repository would promote the use of common tasks and data sets as benchmarks for each subfield. It would lead to the development of shared, widely accepted comparison metrics.

A second, practical, purpose of the repository is to provide a centralized resource that developers can access when they want to start a modeling project. The repository should facilitate finding all the available models for one’s needs and purposes. This includes source code, executable, documentation, papers, and support community. Finding all the existing models for a given task or problem can be difficult since while some tasks are well identified, others can arise in many different forms. The repository would also make available all relevant behavioral or neuroscience data for a given task. The raw data would be provided rather than the aggregate analyses provided in publications to provide additional constraints for the developing of increasingly refined models. Finally, together with data the repository would make available an implemented version of the corresponding experimental tasks. Too much time (as much as half by some estimates) is spent in modeling projects on (re)implementing and connecting to task environments. Often different modelers abstract away from a common task and thus prevent models from being directly comparable.

Thus, the repository would provide an immediate and organized way to access an overview of relevant information, especially key findings of specific subfields for which to develop and validate models. Making available all relevant results would promote broad and integrated rather than partial and selective accounts. Conversely, the repository would provide a consistent and comparable record of activity for the various modeling frameworks. This would provide an archival record of the range of coverage and would highlight the core focus of each framework. It would also encourage keeping models updated to keep credit for successive versions of the framework.

Another function of the repository is to enable the reuse and integration of models. That would in turn promote consistency in parameters across models and discourage excessive (i.e., post hoc) parameter fitting and encourage the adoption of consensus values. Similarly, reuse and integration of models would promote ontological consistency in domain representation. The availability of standard ways of encoding knowledge for specific domains would enable the development of more complex, comprehensive models validated over broader range of findings.

A practical benefit of an integrated repository would be to encourage the development of modeling tools and standards. Developing modeling tools (e.g., for model editing or parameter search) is a rather esoteric niche with little benefits. Making them available to a broad community would benefit the community through improved productivity. The repository would also promote the development of standards, such as for the integration of models and tasks environments. This would raise productivity as well as provide additional constraints on models.

6.2 Practical Issues

While such repositories have proven successful in fields such as Biology and Physics, a major practical issue is how to bootstrap them to the point where they become self-sustaining. A key enabling factor is to give proper credit for uploading materials. This would require limiting submissions to materials associated with published papers, or subjecting submissions to independent peer review. To encourage submission, making models and data available in repository should be made a condition of publication and/or funding (as happens in other fields). This can easily be done for specialized conferences (e.g., NIPS, ICCM, BRIMS). Another incentive is that making behavioral data available for a given task will establish it as a de facto benchmark for its subfield. This will lead to a convergence towards a standardized set of tasks that will keep expanding rather than remain static and thus subject to be gamed, as is often the case with fixed benchmarks.

Tying the repository into an external computational system would allow users to make use of that system with no extra investment in effort. Examples of such external systems include simulation systems (e.g., Unreal Tournament), model running and parameter optimization system (e.g., MindModeling.org) and experiment system (e.g., Eprime). This would also enforce some code-compliance and standardization policies.

Irrespective of popularity, practical issues remain in making a repository successful. Simply uploading tasks and model code is not enough. A number of issues should be considered. Most fundamentally, a standard interface between cognitive models and task environments is needed to assure portability across tasks and models. Tasks and models could only be included in the repository when they are compatible, ensuring interoperability. If both tasks and models comply with the interface, both scientific (principled model comparison, separation between task and model) and technical (reusability, productivity) goals will be enhanced and the exponential growth associated with systems embracing common standards (e.g., the Internet, the Personal Computer) will then be possible. The primary scientific obstacle to such a common interface is to agree upon a common level of description across models. The alternative is to adopt a multi-level model approach that integrates models across multiple grain scales.

Another maintenance issue is how computational models need to be updated and kept current. Developers should have incentives to maintain their code up-to-date to claim cumulative credit from models developed under previous versions of their framework. The main issue is how much standardization should be required (e.g., fixed parameters, common knowledge representation) for a framework to claim an integrated account across the models that it supports.

Most practically, infrastructure funding for the repository should be provided by a central funding source, e.g., Department of Defense (DOD) research agencies, DARPA, NSF, or private foundations. The alternative is incremental funding through individual projects contributing ancillary development to the repository, which would likely result in slower, piecewise development. Even if centralized, the repository developed should be focused on the modeler's needs through informal pools and surveys to make sure that it corresponds to actual developmental patterns and supports the modeling activity.

7.0 DISCUSSION AND CONCLUSION

From a human factors perspective, cognitive architectures and their neural underpinnings provide very strong constraints on our cognitive abilities. In particular, as it relates to our central information processing bottleneck, the conditional routing model has a direct bearing on the modeling of our interaction with complex information environments, such as many military systems and situations, where our performance is primarily determined by our ability to acquire, understand, and process information. Thus integrating our cognitive bottlenecks allows us to model the impact on performance of information-rich environments increasingly common in network-centric warfare. The development of that infastructure is predicated on the assumption that more information afforded by always increasing computing power and ubiquitous sensors is always better. However, the result of more information is not always better performance if it overwhelms our ability to understand and use it effectively in our decisions. As Herbert Simon put it, "Moore's law fixes everything but us".

As mentioned previously, this focus on the inner loop of information processing also enables a precise modeling of the process of interpreting and executing instructions. Instructions feature prominently in a number of military processes, from training to interacting with local populations. Cognitive modeling has often been used to analyze and quantify training processes such as skill acquisition and decay. An equally important application might be to use it to model how instructions are processed by other populations interacting with our military, and determine for instance what the most effective instructions would be to overcome language and knowledge barriers.

When evaluating courses of action, it is essential to be able to quantitatively predict their effect on the various civilian and military stakeholders to which they apply. Paradigms such as IPD2 allows us to develop and validate cognitive models in settings that feature a complex calculus of personal interests, power relationships, and conflicts of interests. Cognitive architectures are not meant as normative tools, but rather can accommodate models designed to reflect individual differences and their impact on decision-making. Those differences can extend from basic cognitive capacity to the interplay between concepts such as emotions and decisions, and to complex representation structures underlying knowledge and culture.

Because so many processes in military operations involve the interaction of multiple, usually large numbers of, cognitive agents, it is essential for our models of cognitive processes to scale up to, and preserve the bulk of their validity in, large-scale applications. This is by no means a given since the dynamics of multiple interacting models are often quite different from those of a single model confronted with a static task and environment. Qualities such as learning, adaptivity, unpredictability, and meta-cognitive flexibility become prominent. Implementing processes possessing those qualities at a tractable computational cost is one of the major challenges in the field of modeling and simulation. Our principled approach, grounded in computational cognitive processes validated at the most basic levels of neuroscience, can still achieve that scalability through the adoption of multiple layers of conceptual and computational abstractions, preserving the core of its fidelity while enabling flexible, tractable integration in a broad range of simulation frameworks.

ACKNOWLEDGEMENTS

The authors were supported by MURI grant FA95500810356 from the Air Force Office of Scientific Research, DTRA contract HDTRA1-09-1-0053 from the Defense Threat Reduction Agency, grant FA9550-08-1-0404 from the Air Force Office of Scientific Research, and contract N00014-07-C-0912 from the Office of Naval Research. The authors would like to thank Harold Hawkins, Jun Zhang, Robin Burke and David Luginbuhl for their support and feedback.

REFERENCES

[1] Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

[2] Anderson, J. R. (1993). *Rules of the Mind.* Hillsdale, NJ: Erlbaum.

[3] Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford, UK: Oxford University Press.

[4] Anderson, J. R. (2009). Using neural imaging to investigate learning in an educational task. Presentation at the *36th Carnegie Symposium on Cognition*, June 3, 2009, Pittsburgh, PA.

[5] Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y . (2004). An integrated theory of the mind. *Psychological Review 111,* (4). 1036-1060.

[6] Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

[7] Anderson, J. R., & Lebiere, C. (2003). The Newell test for a theory of mind. *Behavioral and Brain Science*, *26*, 587-537.

[8] Anderson, J. R., Taatgen, N. A., & Byrne, M. D. (2005). Learning to achieve perfect timesharing: Architectural implications of Hazeltine, Teague, and Ivry (2002). *Journal of Experimental Psychology: Human Perception and Performance, 31(4)*, 749-761.

[9] Christiansen, M. H. and Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31(5)*, 489–509.

[10] Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S., Hau, R., Hertwig, R., Stewart, T., West, R., Lebiere, C. (2010). A choice prediction competition, for choices from experience and from description. *Journal of Behavioral Decision Making 23(1)*: 15-47.

[11] Fay, N., Garrod, S., Roberts, L., and Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science, 34(3)*, 351–386.

[12] Garrod, S., Fay, N., Lee, J., Oberlander, J., and Macleod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science, 31(6)*, 961–987.

[13] Gigerenzer, G., & Todd, P. M. (1999). *Simple Heuristics That Make Us Smart.* Oxford University Press.

[14] Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27*, 591–635.

[15] Kirby, S., Hurford, J., 2002. The emergence of linguistic structure: An overview of the iterated learning model. In: Cangelosi, A., Parisi, D. (Eds.), *Simulating the Evolution of Language*. Springer Verlag, London, Ch. 6, pp. 121–148.

[16] Laird, J. E., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33(1)*, 1-64.

[17] Lebiere, C., & Anderson, J. R. (1993). A connectionist implementation of the ACT-R production system. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 635-640). Mahwah, NJ: Erlbaum.

[18] Lebiere, C., Anderson, J. R., & Reder, L. M. (1994). Error modeling in the ACT-R production system. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, 555-559. Hillsdale, NJ: Erlbaum.

[19] Lebiere, C., Wallach, D., & West, R. L. (2000). A memory-based account of the prisoner's dilemma and other 2x2 games. In *Proceedings of International Conference on Cognitive Modeling*, pp. 185-193. NL: Universal Press.

[20] Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance. 1. Basic mechanisms. *Psychological Review, 104(1)*, 3-65.

[21] Newell, A. N. (1973a). You can't play 20 questions with nature and win: Projective comments on papers in this symposium. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

[22] Newell, A. N. (1973b). Production systems: Models of control structure In W. G. Chase (Ed.), *Visual information processing* (pp. 526-547). New York, NY: Academic Press.

[23] Newell, A. N. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

[24] Rapoport, A., Guyer, M. J., and Gordon, D. G. (1976). *The 2x2 Game*. University of Michigan Press, Ann Arbor, MI.

[25] Reitter, D. (in press). Metacognition and multiple strategies in a cognitive model of online control. *Journal of Artificial General Intelligence*.

[26] Reitter, D., Juvina, I., Stocco, A., & Lebiere, C. (2010). Resistance is futile: Winning lemonade market share through metacognitive reasoning in a three-agent cooperative game. In *Proceedings of the 19th Behavior Representation in Modeling & Simulation (BRIMS)*, Charleston, SC, 2010.

[27] Reitter, D., & Lebiere, C. (2010). Accountable modeling in ACT-UP, a scalable, rapid-prototyping ACT-R implementation. In *Proceedings of the 10th International Conference on Cognitive Modeling*, Philadelphia, PA.

[28] Reitter, D., & Lebiere, C. (in press). Towards explaining the evolution of domain languages with cognitive simulation. *Journal of Cognitive Systems Research*.

[29] Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1*. Cambridge, MA: MIT Press.

[30] Stocco, A., Lebiere, C., & Anderson, J. R. (2010). Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychological Review, 117*, 530-574.

[31] Stocco, A., Lebiere, C., O'Reilly, R. C., & Anderson, J. R. (submitted). The Role of the Basal Ganglia – Anterior Prefrontal Circuit as a Biological Instruction Interpreter. *First Annual Conference on Biologically Inspired Cognitive Architectures (BICA).*

[32] Sun, R. (2007). The motivational and metacognitive control in CLARION. In W. Gray, (Ed.), *Integrated models of cognitive systems* (pp. 63-75). New York: Oxford University Press.

[33] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684): 440–442.