Choosing and getting started with a cognitive architecture to test and use human-machine interfaces

FRANK E. RITTER

Applied Cognitive Science Lab, School of Information Sciences and Technology

The Pennsylvania State University, University Park, PA 16802
frank.ritter@psu.edu

Keywords: cognitive models, cognitive architectures

1. Introduction

This article provides a tutorial review of creating cognitive models with cognitive architectures to help with human-machine interface design. It is becoming increasingly popular and increasingly possible to consider creating cognitive models to assist in design, particularly of users of computer interfaces, and also of human-machine interfaces.

As these models and modeling mature, we will have an approach that can be used to evaluate and test a wide range of interfaces by simulating the human component of the system, and how humans interact with interface and machine components. The requirements for this approach can sketched (Kieras, 2003; Ritter, Van Rooy, & St. Amant, 2002) and point to early examples (Byrne, Wood, Sukaviriya, Foley, & Kieras, 1994; Kieras, Wood, Abotel, & Hornof, 1995).

Commercially prepared cognitive models are already being used for system analysis, although in more limited ways than electrical circuit designers can use their CAD/CAM systems. A few notable examples of this work available as integrated systems include the IGEN system (Zachary, Jones, & Taylor, 2002), the Midas system (Laughery & Corker, 1997), Apex (Freed & Remington, 2000), and the Jack anthropometric system (Badler, Erignic, & Liu, 2002)[1]. The commercial systems that are available do not seem extendible by researchers interested in expanding the behavior and capabilities included.

A prototypical example to illustrate what can be done is a model that examines cell phone design (St. Amant, Horton, & Ritter, 2004). Models were created in ACT-R and GOMS that could perform five tasks on ten cell phone designs. These models performed the five tasks across all of the phone designs. The ACT-R model in particular had access to problem solving and interacted with the cell phone images and representations of an interactive environment. Their results were all basically consistent with the empirical data that was later gathered. The resulting models were then used to optimize the design of the phone interfaces, leading to an estimated time savings of about 30%. Given the widespread use of cell phones, this possible time savings cumulated across all users represents several human lifetimes.

The range of human capabilities that can be considered for inclusion in these models is quite broad and much work remains. The current models, while becoming useful when considering many real design decisions, do not yet include very much of human perception and motor output; they do not have very complex or error tolerant models of error correction and recovery when interacting, and they cannot be routinely applied to a complex interface. There is not a lot of sharing of task models and interfaces. This can be put in contrast with the Lisp, Java, and perhaps expert system shell communities, where contributions are shared on a much more regular basis.

This tutorial review examines how to get started with cognitive models in cognitive architectures with some emphasis on using the resulting models to test interfaces. It notes several cognitive architectures that could be or are being used for evaluating interfaces and predicting task time and (often) errors. The article provides practical comments on how to learn to use a cognitive architecture, providing general guidance as well as pointers to specific resources. The article also examines one of the most vexing questions for modelers, that of how to prove or validate the resulting model. The article concludes by noting some of the most exciting, current problems.

2. How to choose an architecture

You will often start your choice of a cognitive architecture with a problem in mind, and this is important. While in the fullness of time the architectures can be expected to become similar because they are all modeling human cognition, they are forces working to keep them different. Like simulations in other domains, their focus on different domains or levels of analysis may keep them somewhat different.

If you are a new researcher, you may be working where an architecture has already been chosen for you. In that case, you may wish to keep in mind its strengths and limitations, and notice your own new research problems.

2.1 Why use a cognitive architecture?

Before proceeding, it is worth noting what cognitive architectures are, and refer to material to explain them and their approach to modeling. Cognitive architectures are

[1] Note that there is a Jack anthropometric system and a JACK intelligent agent system, same first name, but different approaches and different developers (one a US university, the other an Australian company).

an approach to modeling behavior that assumes that there are two components to behavior, the architecture and knowledge. The architecture is composed of cognitive mechanisms that are fixed across tasks and basically fixed across individuals. These mechanisms typically include some form of perception and motor output, some sort of central processor, some working memory or activation of declarative memory, and some way to store and apply procedures. These mechanisms are used to apply task knowledge to generate behavior.

Newell's (1990) book on unified theories of cognition introduces this approach. ACT-R (Anderson & Lebiere, 1998) probably comes the closest to realizing it currently. Newell's book includes a list of reasons for using a cognitive architecture. To briefly summarize some of the most important, a cognitive architecture proposes that the same mechanisms are used for different tasks, which is parsimonious. The cumulation and unification of results to a central source, are aims of science. When the cognitive architecture is realized as a computer program, it supports these aims by using the architecture itself to serve as a focus for unification. The resulting architecture can then be reused and the effort to create it amortized over multiple projects.

The use of a cognitive architecture also allows for model (knowledge) reuse, but this has been done less than I think Newell anticipated. We are finding that reuse of displays of model behavior (Ritter, Jones, & Baxter, 1998; Tor, Ritter, Haynes, & Cohen, 2004) may perhaps be a more approachable way towards reuse. In the cellphone project (St. Amant et al., 2004), the tasks were reused, but while there were similar models in ACT-R, these had to be created anew here because exact models did not exist.

And finally, the use of a cognitive architecture helps create complete agents opening the way to applications, which is the subject of this article and this special issue. These models can be used to use and thus test interfaces, to serve as opponents or colleagues in synthetic environments, and to run robots.

2.2 Types of architectures

There are several types of architectures that are or that could be used for evaluating interfaces and predicting task time and (often) errors. These include descriptive architectures, symbolic and hybrid architectures, intelligent agent architectures, and connectionist architectures.

The simplest are descriptive architectures like GOMS (John & Kieras, 1996) and the Keystroke-level model (Card, Moran, & Newell, 1983). Models created in these architectures are used to help in system design (e.g., Gray, John, & Atwood, 1993). They are descriptions of behavior rather than generators of behavior for testing interfaces. They can be used to predict the time to do a task, but the actions have to be specified; they do not address problem solving that might be required to do a task. Current work has attempted to make this approach easier to apply to simple interfaces (e.g., Nichols & Ritter, 1995), to automate the more complex behaviors possible (e.g., Freed & Remington, 2000; Matessa, in press), and to unify GOMS and ACT-R creating a higher level language for ACT-R (St. Amant & Ritter, in press). Some of the commercial systems used for interface and system design started with higher level descriptive architectures that computed the time to do larger tasks, but these architectures have tended to migrate towards including information processing and most can now perform the task of interest with an external simulation.

Symbolic and hybrid architectures are most commonly used by researchers in this area. Soar and ACT-R are examples of these two types. Informally, both types can be referred to as cognitive architectures. These architectures support creating knowledge and applying it to situations. The hybrid architectures, ACT-R in particular, tend to provide more action on the level that interface design is currently viewed, that is, reaction times in *ms*, and the possibility of modeling a range of types of errors.

Intelligent agent architectures have been used to explore interfaces (e.g., St. Amant, 2000). They are useful for testing the range of performance of an interface, and for ensuring that an interface can be used to perform a task. They are not designed to make strong predictions about difficulty of use by humans, however. Connectionist architectures have been used extensively in psychology to model behavior, but they have been little used in modeling interaction. They appear to focus on different types of behavior than have been focused on in interface use. They are likely to be useful when modeling the details of perception, and their memory blends and errors are represented to some extent in the hybrid architectures.

2.3 Reviews of architectures

Knowing your potential application will help you choose an appropriate architecture. Architectures have different strengths. Soar, for example, appears to support larger knowledge bases than ACT-R, but does not provide as much support for detailed timing predictions (e.g., Byrne, 2001; Gray & Boehm-Davis, 2000).

There are now several reviews of cognitive architectures that can help you choose an architecture to use. The first review, still helpful although clearly dated, was a special issue of the *SigART Bulletin* (1991). Pew and Mavor's (1998) report is more recent. Their book reviews architectures developed in the US. The architectures they review, such as Soar and ACT-R, are fairly well developed. Further reviews and comparisons of Soar and ACT-R may be helpful, as these are two of the most widely used architectures (Johnson, 1997; Johnson, 1998; Ritter, Shadbolt, Elliman, Young, Gobet, & Baxter, 2003, Appendix B).

Ritter et al.'s (2003) State of the Art Report (as labeled by the publisher, a SOAR report) is an update and extension to Pew and Mavor's report. Their report reviews a set of architectures not included in Pew and Mavor's report, including the Java Agent Construction Kit (JACK: Busetta, Rönnquist, Hodgson, & Lucas, 1999), which is a belief-desires-intentions (BDI) architecture implemented in Java; the COGENT meta-architecture (Cooper & Fox, 1998) that has had some success in teaching; and PSI (Detje, 2000; Dörner, 2003), an architecture that includes physiological drives as a basis of emotions. Ritter et al.'s report also includes a review of current problems and directions for research, which is reviewed in the conclusions. A similar report (i.e., including Silverman, Cornwell, & O'Brien, in press) focused on emotion within architectures will be available shortly.

Even these current books do not include several new and a few old architectures that should be mentioned. Langley (1996) and his research group have an architecture that appears to provide a more schema-based approach. Hybrid architectures, such as Clarion (Neveh & Sun, in press; Sun, Merrill, & Peterson, 1998), attempt to create architectures with sub-symbolic and symbolic representations. Further examples include hybrid versions of blending parts of ACT-R, Soar, and EPIC (e.g., Chong, 2001). These are becoming interesting variants in their own right.

There are a variety of architectures being developed for modeling social agents, that is, models that interact with a few to a large number of other agents (e.g., Carley, 1996; Yen, Yin, Ioerger, Miller, Xu, & Volz, 2001). Presentations at the CASOS conference (e.g., www.casos.cs.cmu.edu/events/conferences/conference_2004.html) often explain advances in this area. These social architectures tend to have less information processing capabilities, but appropriately more communication capabilities as well as including more instrumentation to record and analyze the behavior of groups from 10 to 1,000.

AMBR is a large scale project to compare cognitive architectures based on how they interact with a common task. Their results may help you choose a cognitive architecture. The AMBR project has provided two large scale simulations and had models written in a variety of architectures (Gluck & Pew, 2001a, 2001b; Pew & Gluck, in preparation). In addition to ACT-R and a modified version of Soar, these comparisons have included COGNET/iGEN (a commercial architecture from CHI Systems), and D-COG (an architecture developed by the US Air Force).

This section has described a range of architectures to consider and noted several reviews that provide comparisons. The reader will have to choose their own cross to bear, according to what they want to model and the resources available to them. Two architectures were used in the cell-phone example (St. Amant et al., 2004) for comparison. One (GOMS) was chosen because it is commonly and easily used. The other (ACT-R) was chosen because it is commonly used, it supports problem solving, is extendable, and it will be able to use the results of several related projects.

3. How to learn about architectures and models

It is generally acknowledged that learning how to create cognitive models is not a simple process. There are materials to help with this process, organized here by presentation media; they could be organized through the stages of data gathering to model building and testing as well.

3.1 General textbooks on simulation and modeling

There are some general textbooks on simulation and modeling that would be helpful. Pew and Mavor (1998) implicitly provides some overview. A book in the Sage methodology series (Tabor & Timpone, 1996) explicitly provides just an overview. Books on mathematical psychology (e.g., Greeno, 1968; Townsend & Ashby, 1983; Wickens, 1982) offer some guidance, but as their approach is based on theories that typically have a closed form or with much different, simpler assumptions, they do not always offer much guidance. None-the-less, these books do teach some basic assumptions and lessons that are not yet in a cognitive modeling book.

3.2 Textbooks and materials on cognitive modeling

There are some textbooks that attempt to summarize cognitive modeling and in some cases attempt to teach it. These are worth examining. Boden (1988) provides an overview of several major approaches, including Newell and Simon's approach as well as connectionist approaches.

Those working with connectionist architectures have several books to choose from (McClelland & Rumelhart, 1988; McLeod, Plunkett, & Rolls, 1998; O'Reilley, 2000). These books introduce modeling as well as an associated software package. Through worked examples they provide jumping off places for other projects and for further work.

Cooper (2002) has recently published a book on the COGENT system. Like the connectionist books, it too provides a set of examples, along with comments about model building.

The van Someren, Barnard, and Sandberg book (1994) comes closest to providing a book on how to create cognitive models. It is too brief, but starts to touch on many of the important topics, such as how to gather verbal protocols, how to compute inter-rater reliability, and the use of task analysis.

Ericsson and Simon's (1993) book has to be included here. It provides the rationale for using verbal protocols as data, as well as the many limitations of verbal protocols. It also includes some practical advice, but not nearly enough. Work continues on understanding how to use other non-verbal protocols such as mouse moves (Baccino & Kennedy, 1995), eye-gaze (Anderson, Bothell, & Douglass, 2004), and demasking (Seifert, 2001).

Work on sequential data analysis needs to be included as well. Modelers interested in the sequential predictions of their models would be well advised to become a student of sequential data representations (Sun & Giles, 1998), sequential data analysis (Gottman & Roy, 1990), and exploratory sequential data analysis (Sanderson & Fisher, 1994). There are useful tools in this area to help with coding and analyzing data (e.g., MacShapa: Sanderson, James, & Seidler, 1989, which has been updated since). Reviews of techniques and tools seem to be done every few years (Fielding & Lee, 1991; Ritter, 1993; Sanderson & Fisher, 1994).

3.3 Exemplar books and monographs

There are several books and monographs that are worth studying because they teach by example fairly well, not because they often or very directly give proscriptive advice. Many people have learned this way. Newell and Simon's (1972) *Human problem solving* is probably the canonical one. While few people have read it cover to cover (I think I know of one person, and I have read about two-thirds), it provides numerous examples worked out in great detail, and teaches the ethos and spirit of the approach. Baxter's (1997) report is also presented in this way, and provides a more current example for Soar. Some Soar and ACT-R theses also provide examples (e.g., Wiesmeyer, 1992), but often do not provide much help for those interested in learning the process.

There are now several edited books on cognitive modeling (Polk & Seifert, 2002; Rosenbloom, Laird, & Newell, 1992). These do not provide a unified treatment, but do provide numerous lessons and further examples. Simon's (1979, 1989) books of collected works in this area (the checkerboard books) have numerous examples, many of which are still worth building on, and all are worth learning.

3.4 Useful articles

Yost and Newell's (1989) article is helpful, as it attempts to explain the process of building a model, but it is tied to a single architecture and should be more widely read. There are other useful papers, but they are short and do not provide the full story. Interested students of modeling will find them helpful (Kieras, 1985; Ritter & Larkin, 1994; Sun & Ling, 1998). There are numerous examples of models and their fit to the data, but they tend not to explain the steps of model creation from the data – the model tends to appear in final form as Athena did from the head of Zeus.

3.5 Programming style and informal mechanisms

Users of architectures have found that in addition to the architecture and knowledge there is another component that needs to be formalized. Because models within the architectures are theories, how they use the mechanisms – in a uniform or ad hoc basis – are part of the theory.

Newell (1990) noted this when he said that there is more in your architecture than you would expect. Learning how to use an architecture and how not to misuse it has to be learned, not just individually, but as a community. Kieras, Wood, and Meyer (1997) referred to it as rules for creating models. Kieras, 2003 provides an update to this. Newell (1990) and the Psychological Soar Tutorial (http://acs.ist.psu.edu/nottingham/pst/pst-ftp.html) refer to it as listening to the architecture and using it appropriately. Some in the Soar group have recently started to formalize how to program Soar in a document where the how-to is referred to as dogma (Nuxoll & Laird, 2003). Similar requirements are already apparent as we create the COJACK architecture (Norling & Ritter, in press).

What all of these authors are referring to is a set of conventions that are adhered to when creating or programming the model. Examples of these conventions for Soar include using only one value per attribute in Soar, not putting too much information in a single state, and not creating operators that are overly complex.

Learning this architectural component is difficult because it is not yet formalized (although model compilers offer the promise of doing this, e.g., St. Amant & Ritter, in press). It is perhaps this type of knowledge that is missing when modelers have difficulties or give up. The old rule of thumb was that you had to visit an established site to absorb this information, and while numerous steps have been taken in recent years to reduce this requirement, such visits are still a good idea.

3.6 Conferences, tutorials, and online materials

Conferences (e.g., the Cognitive Science Conference, International Conference on Cognitive Modeling) and workshops (e.g., the Soar, ACT-R, and Cogent Workshops) offer opportunities to learn current programming (modeling) paradigms and to meet other modelers. Such paradigms have been actively debated in panels as Soar and ACT-R workshops.

There are also more formal places to learn. Tutorials are now often offered at the relevant conferences, and these can serve as useful introductions. The ACT-R summer School and the German Cognitive Autumn School (Herbstschule Kognitionswissenschaft), while only available occasionally, offer excellent opportunities to get started and to learn more.

Finally, there are now online materials. On their main web sites both ACT-R (http://act.psy.cmu.edu/) and Soar (http://sitemaker.umich.edu/soar) now have online tutorials (you print them and do the exercises), and they both have Frequently asked questions lists (http://acs.ist.psu.edu/act-r-faq, http://acs.ist.psu.edu/soar-faq). Other architectures are likely to have the same, if not now, they will have them soon as the standard of support rises for new and existing users. These tutorials and FAQs also raise issues that most architectures will wish to address, so they should provide value even for users of other architectures.

3.7 Psychology and computer science materials

There are two further areas important for the success of modeling to keep in mind: psychology (the data to be modeled), and computer science (the tools used in modeling). If you are a computer scientist coming to cognitive modeling, you will need an overview of the information processing view of psychology. Anderson's books on psychology (1996) and on learning and memory (1995) are excellent introductions and good overviews. (There are others as well.) If you are interested in more specific areas, textbooks in those areas will also be helpful.

If you wish to have access to a wide range of data useful for modeling, either to help build an architecture or else to provide additional data to extend the coverage of the model, engineering psychology can provide this. Wickens' text books (Wickens, Gordon, & Liu, 1998; Wickens & Hollands, 2000) provide an overview, and the Engineering Data Compendium (Boff, Kaufman, & Thomas, 1986; Boff & Lincoln, 1988) provides a detailed view that is sometimes helpful.

It was first noted by Kieras (1985) that it is very useful, perhaps even essential, that modelers know the language underlying their architecture to assist in modeling. After the model is built, additional apparatus will have to be built to include running the model multiple times (if it is stochastic), to explore variants of the model, to run the model on a variety of tasks, and to provide the model access to a task simulation. For current modelers, this can mean studying Lisp, Tcl/Tk, and Java. Online resources and summaries of learning materials for these languages can be found, for example, in the frequently asked questions lists for the architecture (e.g., http://acs.ist.psu.edu/act-r-faq, http://acs.ist.psu.edu/soar-faq). Psychologists who just set themselves the task to learn ACT-R and not Lisp will run into difficulties, and will either learn Lisp or quit.

4. How to test your model[2]

Modelling is always a purpose driven act. Thus every model has a purpose (or set of purposes). Testing and validation has to be done with the purpose of the model in mind. For science, the role of the proposed mechanisms to account for behavior is pretty common. Thus testing is probably the better label for this step. For engineering and design, the usefulness and usability of the model are being considered, so validation is probably a better label for this step.

[2] An earlier version of this section was presented at a Symposium on Model Fitting and Parameter Estimation at the ACT-R Workshop, 2003.

How to test and validate your model is a problem that has vexed many researchers. I have met several people and I have spent time myself looking for a statistic to prove cognitive models. This is a fruitless search that I suspect is repeated far too often. There is an approach to testing models that I think is productive. Campbell and Bolton (in preparation) provide a longer explanation that will be useful to those particularly interested in this topic. After the theoretical background is provided, some practical suggestions are provided.

4.1 Theoretical background

Grant's (1962) paper on the strategy and tactics of investigating models argued that there were two important aspects for testing a model, that (a) the model was worth taking seriously, and (b) you could see where the model was wrong so that you could improve it. This is consistent with Newell's (1990, p. 507) view of how to develop unified theories of cognition (UTC): what is the current bar (standards), does this theory (or model) raise it, and what are the further regularities to be included in the future? I like this approach as it lets me make progress, or at least be happy. I have seen others trying to prove their model, and they are not and cannot be happy because proving your model is equivalent to accepting the null hypothesis (see any elementary statistics book for a description of the dangers of that).

Taking a model seriously depends on what other models are available and what you want to do with it. What is the current best model? If you look at current theory/data comparisons of task performance models (e.g., ACT-R), the models can typically match a single type of data or a few kinds of data on a single task. For example, the data compared with the model will include one or a few of reaction time means, the sequence of task actions, groupings of task actions into strategies, error rates and types, and trends and variance in all of these. Few models have had their predictions compared to all of these aspects of data. Fewer yet have been compared to data from multiple tasks. Gobet and Ritter (2000) describe this approach; Lovett, Daily, and Reder (2000) independently have provided an example.

Sometimes in a new area of modeling it will be enough report the performance of model, that is, that it can do the task (currently models of teamwork and emotions seem to use this approach). More advanced models may report the correlation between the model's predictions and data, which Simon and Grant both recommend. Correlations currently appear to be a good standard, and they often lead me to take models seriously.

Summarizing the match across these sets of regularities can be done in multiple ways. For example, John (1996) has used a type of bar chart across a set of different types of behavior being matched. Further details of this approach are presented in Newell's (1990) book, and briefly expanded in Ritter (1993) as criterion-based cognitive modeling. This approach, of criterion-based cognitive modeling, is a way to protect models because it defines the range and performance expected from a given model. Schunn and Wallach, (2001) also provide many useful suggestions.

With multiple types of data with multiple values and multiple displays, how can one compare theories? I currently think that Grant's question, "is a theory worth taking seriously?", can be seen at least partly as a social process. Theories will correspond to the data on a number of dimensions. Reducing their fit to a single number for comparison to choose the best model is likely to be difficult when complex models or complex data sets are considered. Some models are admirable because they do not touch the simulation and offer new worlds to models (but have an unreported or poor fit). Another model may be interesting because it opens up new areas of data to be included into ACT-R or Soar. A third model may be interesting because it shows how to use a genetic algorithm to fairly test a wide range of ways to adjust models to fit a dataset (e.g., what develops in children?, Tor & Ritter, in press).

In each case, the judgment of "is this model interesting" is based on other models, how well the model fits the data, how applicable the theory, how easy the theory is to use, and a host of other factors. Estimates of future applicability is also important. "Science, like politics, is the art of the possible", said Newell, and I rather strongly agree. That means that I take models that I can download and include with my model much more seriously than those that I cannot inspect or that cost $1,000. (Something interesting is going on here, because except for computer proofs in mathematics, rarely in science are theories cast so strongly as programs, it seems; and estimates of future usability are likely to be inaccurate.)

A recent set of comments (Roberts & Pashler, 2000, 2002; Rodgers & Rowe, 2002) argue that a reader needs to know more about the model predictions to data comparison than just the fit. They argue that readers need to know what kind of data that the theory cannot fit, the variability of the data, and the likelihood of fitting data. Roberts and Pashler's stance appears to be consistent basically with Grant's two step process, but they ask for more details. The details they ask for appear to me to be more relevant for simple models covering well trod but narrow ground rather than broad, approximate theories that current cognitive models often look like. Roberts and Pashler do request a standard that is worth striving for, but they also appear to be overly harsh. Newell (1990) argued for allowing models time to develop (citing Hebb, "don't bit my finger look where I'm pointing"), and to allow them to have success in multiple ways. A model that performed a new type of learning or problem solving would be inappropriately rejected by Robert's and Pashler's criteria.

Roberts and Pashler prefer theories that predict surprising data. I also find much of psychology data surprising, which they do not, and thus I think predicting this data is worthwhile. I know of several theories that do not predict smooth curves and the data matches these non-smooth curves. Finally, I believe that task performance is much more important than fits to data because task performance is a prerequisite for generating behavior and thus for more autonomous predictions and applications.

4.2 Practical recommendations

So, (a) I recommend that you tell us about your model's predictions, what the data look like, and how the model's predictions correspond to the data in detail, enough so that we can see that the model is worth taking seriously. Because the judgement is based on other alternative models (if any), there is no a priori quality required. There is not a value of r^2 that must be satisfied, although the r^2 of competing models are a good yardstick. You might also note a model's other virtues, such as ease of use, and consistency but not yet correlation with large swaths of behavior.

There are also reasons to dismiss a model. If the model would fit any data, then it is not worth taking seriously (but only if such data already exist, hypothesized data need not apply). If I cannot understand the model; if it is a hack; or if I believe it will not generalize to other data; and I would add now, if it is not part of a UTC, I am less interested.

I also recommend that you (b) Note where the model can be improved. This does not mean including in your paper a laundry list of data that your model does not yet cover because you ran out of time. If you are a reviewer, it certainly does not mean providing 40 pages of comments of places where the authors could extend their model.[3]

Therefore, include in your reports just enough detail on your model's limitations for readers to know that you know where the remaining problems are, and to indicate that you know enough to improve your model, but not to apologize for tasks it cannot yet do. Thus, for the cell-phone model (St. Amant et al., 2004) both processes were done. The model was tested to find out how it compared to existing scientific models in the area. The models are shown to predict the data using a table and a figure to show the correlation. We also noted where it could be improved in the near term. The model was also validated, in that the results showed that the models were usable and made not only accurate but useful predictions. The effects of redesign could not only be predicted, but indicated that redesign could be very beneficial.

5. How to choose a problem to work on

With an architecture in hand you might then wish to choose an interesting and timely problem to work on. In reality, the problem in front of you is likely to have driven you to desire to work with a cognitive architecture to start with, or may have arisen from your use of a particular cognitive architecture. Thus, this section is not usually the last section in your journey, but it is include it at the end as a summary of what I believe are some of the current areas of interest when using cognitive architectures. Other interesting problems and reviews exist in this area, and this section does not describe, of course, all possible problems.

Newell in his *Desires and Diversions* talk (Newell, 1991)[4] emphasized the need to work on the most important problem. While I have colleagues who disagree with me on the necessity of this directed approach, what progress I feel seems to come in the same way as Newell described it, as returning to the same problems and staying focused on a line of research as much as possible; working on an important problem; and working on a problem where you have some comparative advantage due to education or access to resources or affinity. Thus, in introducing these areas as interesting, you may note that along with a variety of colleagues I am working on some aspect of several of these problems.

[3] These two suggestions are consonant with two comments taken from a document a colleague recently was kind enough to share with me, *Levy's Ten Laws of the Disillusionment of the True Liberal*, findable online with a search engine. That is, Law 4b: Good intentions are far more difficult to cope with than malicious behavior; and Law 8: No amount of genius can overcome a preoccupation with detail.

[4] This talk is available online. The recommended path is to go to http://www.ul.cs.cmu.edu/ (Universal Library); click on Multimedia and Lectures, click on Distinguished lectures, click on 1991, and then click on the link to the talk. If you need a hard link, this path, which varies based on machine type and operating system, sometimes resolves to: http://doi.library.cmu.edu/10.1184/LOCAL/4205 . The slides are available through http://diva.library.cmu.edu/Newell/

Pew and Mavor (1998) in the course of their book describe worthwhile projects at the end of each applicable chapter. The problems they note are interesting problems and worth working on. Their descriptions tend to refer to whole areas and to be the size of a multi-year research proposal that might include multiple investigators

Ritter et al.'s (2003) State of the Art report provides a listing of about 23 smaller projects. As a group, these projects are more approachable as PhD, MSc, or class projects in advanced AI, simulation, or modeling classes. These projects are grouped into three main categories, that of extending the coverage of architectures and models, of improving their integration with tasks, and making the models and architectures more usable.

5.1 Projects extending architectures and models

The first project area is to more accurately model human behavior. To highlight just a few of the projects there, one project suggests that including errors in performance will be important. This is of interest to interface designers in safety-critical systems (e.g., Freed & Remington, 1998). Several projects consider learning. The wide varieties of change that learning entails is a large, broad area for work and will be for a long time.

Another interesting project area is including models of emotions, changes of motivation, and changes within and across individuals. Examples of this work are available in my research group (Belavkin & Ritter, 2003; Norling & Ritter, in press; Ritter, Avraamides, & Councill, 2002) and elsewhere (Gratch, in press; Hudlicka & McNeese, 2002; Silverman, Cornwell, & O'Brien, in press).

5.2 Projects improving the integration of architectures with tasks

The second project area is integrating models with other systems, broadly defined. In particular, the most interesting problem to me is providing models with access to interfaces in ways that approximate the richness of human perceptual-motor capabilities as well as including the limitations of human capabilities. Providing models access to tasks has been a constant problem for modelers, of how to provide their models access to the task of interest or an interesting task where the phenomenon of interest can be studied. The field has started with providing the task in the modeling language. When that approach became unwieldy, providing the models access to raw information through sockets in a complete domain. There are good reasons to use a micro-world, a simplified simulation of a larger task (Gray, 2002). These approaches have not been entirely satisfactory, and I have seen many projects flounder on this "uninteresting" technical subtask.

In other sciences instrumentation and essential technology support that caused researchers to fail would make the failure point an interesting problem to those disciplines. Thus I think interaction is an interesting problem for cognitive modeling. In the last few years we have been working to create simulated eyes and hands (Lonsdale & Ritter, 2000; Norling & Ritter, 2001; Ritter, Baxter, Jones, & Young, 2000). In particular, we have been working with St. Amant and his students to create a simulated eye and hand that do not need to instrument an interface to interact with it (St. Amant, Horton, & Ritter, 2004; St. Amant & Riedl, 2001).

5.3 Making the models and architectures more usable

The third and final project area is to improve the usability of the resulting models. If the models are too difficult to build or too difficult to understand, use, or apply, then they will not be used. This is an interesting human-computer interaction design task, of creating and explaining intelligent behavior. It is similar in many ways to expert system development, except the systems are not only required to be intelligent, but intelligent like humans.

We have started to gather descriptions of what users want (Councill, Haynes, & Ritter, 2003), and are working on higher level programming languages (St. Amant & Ritter, in press) and displays (Tor, Ritter, Haynes, & Cohen, 2004). Others are working on this problem as well (e.g., Crossman, Jones, Lebiere, & Wray, 2004; also see the upcoming AAAI 2004 Workshop on Intelligent Agent Architectures: Combining the Strengths of Software Engineering and Cognitive Systems).

5.4 Concluding remarks

It is an exciting time for creating and using cognitive models. The technology continues to mature and the science that can be addressed continues to provide interesting problems. We are a long way from being able to routinely create and apply cognitive models in the way that ANOVAs and regressions can be, but the path forward continues to seem clearer and broader with the passage of time. This article may help you on this path, but like nearly all science, progress will be faster with a mentor and being part of a community.

6. Acknowledgements

As this article is about teaching and learning, I would like to take this opportunity to thank my teachers, from the early science ones, to the one who took me aside to teach me logarithms after school, to Newell and Simon who taught by example. Many of my colleagues have taught me in this area. Discussions with Richard Young have been particularly helpful. My students have also had to teach me how to teach this material; I hope this article helps them and makes their learning easier. Rich Carlson, Martin C. Kindsmueller, Emma Norling, Bill Stevenson, and two anonymous reviewers have provided useful comments. Cindy Carroll and the CMU Library's Information Technology staff who created the Newell video reference. Preparation of this report was partially supported ONR, grants N00014-03-1-0248 and N00014-02-1-0021, and partially funded by the Director of Technology Development, Ministry of Defence, Metropole Building, Northumberland Ave, London WC2N 5BP and was carried out under the terms of Contract No RT/COM/3/006.

7. References

Anderson, J. R. (1995). *Learning and memory*. New York, NY: John Wiley and Sons.

Anderson, J. R. (1996). *Cognitive psychology and its implications* (3rd ed.). New York, NY: W. H. Freeman.

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes. *Psychological Science, 15*(4), 225-231.

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

Baccino, T., & Kennedy, A. (1995). MICELAB: Spatial processing of mouse movement in Turbo-Pascal. *Behavior Research Methods, Instruments, & Computers, 27*(1), 76-78.

Badler, N. I., Erignac, C. A., & Liu, Y. (2002). Virtual humans for validating maintenance procedures. *Communications of the ACM, 45*(7), 56-63.

Baxter, G. D. (1997). *From soup to nuts: Developing a Soar cognitive model of the electronic warfare officer's task*. Working paper No. WP/R3BAIA005/014. Cognitive Modelling unit. Nottingham, UK: Psychology Department, University of Nottingham. http://acs.ist.psu.edu/nottingham/papers/baxter97.pdf [link checked 4 June 2004].

Belavkin, R. V., & Ritter, F. E. (2003). The use of entropy for analysis and control of cognitive models. In F. Detje, D. Dörner, & H. Schaub (Eds.), *Proceedings of the Fifth International Conference on Cognitive Modeling*. 21-26. Bamberg, Germany: Universitäts-Verlag Bamberg.

Boden, M. A. (1988). *Computer models of mind*. Cambridge, UK: Cambridge U. Press.

Boff, K. R., Kaufman, L., & Thomas, J. P. (Eds.). (1986). *Handbook of perception and human performance*. New York: John Wiley & Sons.

Boff, K. R., & Lincoln, J. E. (Eds.). (1988). *Engineering data compendium*. Wright-Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory.

Burton, R. (1998). Validating and docking: An overview, summary and challenge. In M. Prietula, K. Carley, & L. Gasser (Eds.), *Dynamics of organizations*. 215-228. Menlo Park, CA: AAAI.

Busetta, P., Rönnquist, R., Hodgson, A., & Lucas, A. (1999). JACK intelligent agents - Components for intelligent agents in JAVA. *AgentLink News Letter*, 2(Jan.), http://www.agent-software.com [link checked 4 June 2004].

Byrne, M. D. (2001). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies, 55*, 41-84.

Byrne, M. D., Wood, S. D., Sukaviriya, P., Foley, J. D., & Kieras, D. E. (1994). Automating interface evaluation. In *Proceedings of the CHI‘94 Conference on Human Factors in Computer Systems*. 232-237. New York, NY: ACM.

Campbell, G. E., & Bolton, A. E. (in preparation). Implications for model assessment and validation. In R. W. Pew & K. A. Gluck (Eds.), *Modeling human behavior with integrated cognitive architectures: Comparison, evaluation, and validation.* Mahwah, NJ: Lawrence Erlbaum.

Card, S., Moran, T., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum.

Carley, K. M. (1996). A comparison of artificial and human organizations. *Journal of Economic Behavior & Organization, 31*, 175-191.

Chong, R. S. (2001). Low-level behavioral modeling and the HLA: An EPIC-Soar model of an enroute air-traffic control task. In *Proceedings of the 10th Computer Generated Forces and Behavioral Representation Conference*. 27-35. 10TH-CGF-072. Orlando, FL: Division of Continuing Education, University of Central Florida. http://www.sisostds.org/cgf-br/10th/ [link checked 4 June 2004].

Cooper, R. P. (2002). *Modelling high-level cognitive processes*. Mahwah, NJ: Lawrence Erlbaum.

Cooper, R., & Fox, J. (1998). COGENT: A visual design environment for cognitive modelling. *Behavior Research Methods, Instruments, and Computers, 30*, 553-564.

Crossman, J., Wray, R., Jones, R., & Lebiere, C. (2004). A high level symbolic representation for behavior modeling. In *Proceedings of the 13th Conference on Behavior Representation in Modeling and Simulation*. 04-BRIMS-051. http://www.sisostds.org/cgf-br/04Brims/ [link checked 4 June 2004]. Orlando, FL: U. of Central Florida.

Councill, I. G., Haynes, S. R., & Ritter, F. E. (2003). Explaining Soar: Analysis of existing tools and user information requirements. In F. Detje, D. Dörner, & H. Schaub (Eds.), *Proceedings of the Fifth International Conference on Cognitive Modeling*. 63-68. Bamberg, Germany: Universitäts-Verlag Bamberg.

Detje, F. (2000). Comparison of the PSI-theory with human behaviour in a complex task. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modelling*. 86-93. Veenendaal, The Netherlands: Universal Press.

Deutsch, S. E. (1998). Interdisciplinary foundations for multiple-task human performance modeling in OMAR. In *Proceedings of the Twentieth Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Dörner, D. (2003). The mathematics of emotions. In F. Detje, D. Dörner, & H. Schaub (Eds.), *Proceedings of the Fifth International Conference on Cognitive Modelling*. 75-80. Bamberg, Germany: Universitäts-Verlag Bamberg.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.

Fielding, N. G., & Lee, R. M. (Eds.). (1991). *Using computers in qualitative research*. London & Beverly Hills, CA: Sage.

Freed, M., & Remington, R. (1998). A conceptual framework for predicting error in complex human-machine environments. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society*. 356-361. Mahwah, NJ: Lawrence Erlbaum.

Freed, M., & Remington, R. (2000). Making human-machine system simulation a practical engineering tool: An APEX overview. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the 3rd International Conference on Cognitive Modelling*. 110-117. Veenendaal, The Netherlands: Universal Press.

Gluck, K. A., & Pew, R. W. (2001a). Lessons learned and future directions for the AMBR model comparison project. In *Proceedings of the 10th Computer Generated Forces and Behavioral Representation Conference*. 10TH-CGF-067. 113-121. Orlando, FL: Division of Continuing Education, University of Central Florida. http://www.sisostds.org/cgf-br/10th/ [link checked 4 June 2004].

Gluck, K. A., & Pew, R. W. (2001b). Overview of the agent-based modeling and behavior representation (AMBR) model comparison project. In *Proceedings of the 10th Computer Generated Forces and Behavioral Representation Conference*. 10TH-CGF-066. 3-6. Orlando, FL: Division of Continuing Education, University of Central Florida. http://www.sisostds.org/cgf-br/10th/ [link checked 4 June 2004].

Gobet, F., & Ritter, F. E. (2000). Individual Data Analysis and Unified Theories of Cognition: A methodological proposal. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the 3rd International Conference on Cognitive Modelling*. 150-157. Veenendaal, The Netherlands: Universal Press.

Gottman, J. M., & Roy, A. K. (1990). *Sequential analysis: A guide for behavioral researchers*. Cambridge, UK: Cambridge University Press.

Grant, D. A. (1962). Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review, 69*(1), 54-61.

Gratch, J. (in press). A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research*.

Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and microworlds in basic and applied cognitive research. *Cognitive Science Quarterly, 2*(2), 205-227.

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied, 6*(4), 322-335.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance. *Human-Computer Interaction, 8*(3), 237-309.

Greeno, J. G. (1968). *Elementary theoretical psychology*. Reading, MA: Addison-Wesley.

Hudlicka, E., & McNeese, M. D. (2002). User affective and belief states: Assessment and user interface adaptation. *Journal of User Modeling and User Adapted Interaction, 12*, 1-47.

John, B. E. (1996). TYPIST: A theory of performance in skilled typing. *Human Computer Interaction, 11*(4), 321-355.

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction, 3*(4), 320-351.

Johnson, T. R. (1997). Control in ACT-R and Soar. In M. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*. 343-348. Hillsdale, NJ: Lawrence Erlbaum.

Johnson, T. R. (1998). A comparison of ACT-R and Soar. In U. Schmid, J. Krems, & F. Wysotzki (Eds.), *Mind modeling - A cognitive science approach to reasoning, learning and discovery*. 17-38. Lengerich, Germany: Pabst Scientific Publishing.

Kieras, D. E. (1985). The why, when, and how of cognitive simulation. *Behavior Research Methods, Instrumentation, and Computers, 17*, 279-285.

Kieras, D. (2003). Model-based evaluation. In J. Jacko & A. Sears (Eds.), *Handbook for human-computer interaction.* 1139-1151. Mahwah, NJ: Lawrence Erlbaum.

Kieras, D. E., Wood, S. D., Abotel, K., & Hornof, A. (1995). GLEAN: A computer-based tool for rapid GOMS model usability evaluation of user interface designs. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST'95)*. 91-100. New York, NY: ACM.

Kieras, D. E., Wood, S. D., & Meyer, D. E. (1997). Predictive engineering models based on the EPIC architecture for a multimodal high-performance human-computer interaction task. *Transactions on Computer-Human Interaction, 4*(3), 230-275.

Langley, P. (1996). An abstract computational model of learning selective sensing skills. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. 385-390. Hillsdale, NJ: Lawrence Erlbaum.

Laughery, K. R., & Corker, K. M. (1997). Computer modeling and simulation of human/system performance. In G. Salvendy (Ed.), *Handbook of human factors*. New York, NY: John Wiley and Sons.

Lehman, J. F., Laird, J. E., & Rosenbloom, P. S. (1996). A gentle introduction to Soar, an architecture for human cognition. In S. Sternberg & D. Scarborough (Eds.), *Invitation to cognitive science, vol. 4*. Cambridge, MA: MIT Press.

Lonsdale, P. R., & Ritter, F. E. (2000). Soar/Tcl-PM: Extending the Soar architecture to include a widely applicable virtual eye and hand. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the 3rd International Conference on Cognitive Modelling*. 202-209. Veenendaal, The Netherlands: Universal Press.

Lovett, M. C., Daily, L. Z., & Reder, L. M. (2000). A source activation theory of working memory: Cross-task prediction of performance in ACT-R. *Journal of Cognitive Systems Research, 1*, 99-118.

Matessa, M. (in press). Anticipatory eye movements in interleaving templates of human behavior. In *Proceedings of the International Conference on Cognitive Modeling*. . Mahwah, NJ: Lawrence Erlbaum.

McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. Cambridge, MA: MIT Press.

McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes*. Oxford, UK: Oxford University Press.

Naveh, I., & Sun, R. (in press). Simulating a simple case of organizational decision making. In R. Sun (Ed.), *Cognition and multi-agent interaction: From cognitive modeling to social simulation.* New York, NY: Cambridge U. Press.

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

Newell, A. (1991). *Desires and diversions: Carnegie-Mellon University School of Computer Science Distinguished Lecture*. Palo Alto, CA: University Communications. 64 min. video.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nichols, S., & Ritter, F. E. (1995). A theoretically motivated tool for automatically generating command aliases. In *Proceedings of the CHI'95 Conference on Human Factors in Computer Systems*. 393-400. New York, NY: ACM.

Norling, E., & Ritter, F. E. (2001). Embodying the JACK agent architecture. In M. Stumptner, D. Corbett, & M. Brooks (Eds.), *AI 2001: Advances in Artificial Intelligence. Proceedings of the 14th Australian Joint Conference on Artificial Intelligence*. 368-366. Berlin: Springer.

Norling, E., & Ritter, F. E. (in press). A parameter set to support psychologically plausible variability in agent-based human modelling. To appear in *Proceedings of The Third International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS04)*.

Nuxoll, A., & Laird, J. (2003). Soar Design Dogma. Version 0.6, http://ai.eecs.umich .edu/soar/sitemaker/docs/misc/dogma.pdf [link checked 4 June 2004].

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: The MIT Press.

Pew, R. W., & Gluck, K. A. (Eds.). (in preparation). *Modeling human behavior with integrated cognitive architectures: Comparison, evaluation, and validation.* Mahwah, NJ: Lawrence Erlbaum.

Pew, R. W., & Mavor, A. S. (Eds.). (1998). *Modeling human and organizational behavior: Application to military simulations*. Washington, DC: National Academy Press. http://books.nap.edu/catalog/6173.html [link checked 4 June 2004].

Polk, T. A., & Seifert, C. M. (2002). *Cognitive modeling*. Cambridge, MA: MIT Press.

Ritter, F. E. (1993). *TBPA: A methodology and software environment for testing process models' sequential predictions with protocols*. Technical Report No. CMU-CS-93-101. Pittsburgh, PA: School of Computer Science, Carnegie-Mellon University. http://reports-archive.adm.cs.cmu.edu/anon/1994/CMU-CS-94-102.ps [link checked 4 June 2004].

Ritter, F. E., Avraamides, M., & Councill, I. G. (2002). An approach for accurately modeling the effects of behavior moderators. In *Proceedings of the 11th Computer Generated Forces Conference*. 29-40, 02-CGF-002. Orlando, FL: U. of Central Florida.

Ritter, F. E., Baxter, G. D., Jones, G., & Young, R. M. (2000). Supporting cognitive models as users. *ACM Transactions on Computer-Human Interaction, 7*(2), 141-173.

Ritter, F. E., Jones, R. M., & Baxter, G. D. (1998). Reusable models and graphical interfaces: Realising the potential of a unified theory of cognition. In U. Schmid, J. Krems, & F. Wysotzki (Eds.), *Mind modeling - A cognitive science approach to reasoning, learning and discovery*. 83-109. Lengerich, Germany: Pabst Scientific Publishing.

Ritter, F. E., & Larkin, J. H. (1994). Using process models to summarize sequences of human actions. *Human-Computer Interaction, 9*(3), 345-383.

Ritter, F. E., Shadbolt, N. R., Elliman, D., Young, R., Gobet, F., & Baxter, G. D. (2003). *Techniques for modeling human performance in synthetic environments: A supplementary review*. Wright-Patterson Air Force Base, OH: Human Systems Information Analysis Center (HSIAC), formerly known as the Crew System Ergonomics Information Analysis Center (CSERIAC). http://iac.dtic.mil/hsiac/S-docs/SOAR-Jun03.pdf [link checked 4 June 2004].

Ritter, F. E., Van Rooy, D., & St. Amant, R. (2002). A user modeling design tool based on a cognitive architecture for comparing interfaces. In C. Kolski & J. Vanderdonckt (Eds.), *Computer-Aided Design of User Interfaces III, Proceedings of the 4th International Conference on Computer-Aided Design of User Interfaces CADUI'2002*. 111-118. Dordrecht, NL: Kluwer Academics Publisher.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*(2), 358-367.

Roberts, S., & Pashler, H. (2002). Reply to Rodgers and Rowe (2002). *Psychological Review, 109*(3), 605-607.

Rodgers, J. L., & Rowe, D. C. (2002). Theory development should begin (but not end) with good empirical fits: A comment on Roberts and Pashler (2000). *Psychological Review, 109*(3), 599-604.

Rosenbloom, P. S., Laird, J. E., & Newell, A. (1992). *The Soar Papers: Research on integrated intelligence*. Cambridge, MA: MIT Press.

Sanderson, P. M., & Fisher, C. A. (1994). Exploratory sequential data analysis: Foundations. *Human-Computer Interaction, 9*(3&4), 251-317.

Sanderson, P. M., James, J. M., & Seidler, K. S. (1989). SHAPA: an interactive software environment for protocol analysis. *Ergonomics, 32*(11), 1271-1302.

Schunn, C. D., & Wallach, D. (2001). Evaluating goodness-of-fit in comparison of models to data. Online manuscript. Retrieved 1 April 2004 from www.lrdc.pitt.edu/schunn/gof/GOF.doc.

Seifert, K. (2001). Comparison of methods for studying air traffic controller's information strategies. *MMI-Interaktiv, 5, Juli 2001*, http://www.useworld.net /servlet/handlearticle?obj_id=573&cat_id=40&vie=true [link checked 4 June 2004].

SIGArt (1991). Special section on integrated cognitive architectures. *Sigart Bulletin, 2*(4).

Silverman, B. G., Cornwell, J., & O'Brien, K. (in press). Human performance simulation. In J. W. Ness, D. R. Ritzer, & V. Tepe (Eds.), *The science and simulation of human performance.* Amsterdam: Elsevier. http://www.seas.upenn.edu/~barryg/Ch9-final.pdf [link checked 4 June 2004].

Simon, H. A. (1979). *Models of thought*. New Haven, CT: Yale University Press.

Simon, H. A. (1989). *Models of thought, Volume II*. New Haven, CT: Yale University Press.

St. Amant, R. (2000). Interface agents as surrogate users. *intelligence, 11*(Summer 2000), 29-38.

St. Amant, R., Horton, T. E., & Ritter, F. E. (2004). Model-based evaluation of cell phone menu interaction. In *Proceedings of the CHI'04 Conference on Human Factors in Computer Systems*. 343-350. New York, NY: ACM.

St. Amant, R., & Riedl, M. O. (2001). A perception/action substrate for cognitive modeling in HCI. *International Journal of Human-Computer Studies, 55*, 15-39.

St. Amant, R., & Ritter, F. E. (in press). Automated GOMS to ACT-R model generation. In *submitted to the International Conference on Cognitive Modeling*. . Mahwah, NJ: Lawrence Erlbaum.

Sun, R., & Giles, C. L. (1998). *Sequence learning*. Berlin, Germany: Springer.

Sun, R., & Ling, C. X. (1998). Computational cognitive modeling, the source of power, and other related issues. *AI Magazine, 19*(2), 113-120.

Sun, R., Merrill, E., & Peterson, T. (1998). Skill learning using a bottom-up hybrid model. In F. E. Ritter & R. M. Young (Eds.), *Proceedings of the 2nd European Conference on Cognitive Modelling*. 23-29. Thrumpton, Nottingham: Nottingham University Press.

Tabor, C. S., & Timpone, R. J. (1996). *Computational modelling*. Thousand Oaks, CA: Sage.

Tor, K., & Ritter, F. E. (in press). Using a genetic algorithm to optimize the fit of cognitive models. In *Proceedings of the Sixth International Conference on Cognitive Modeling*. . Mahwah, NJ: Lawrence Erlbaum.

Tor, K., Ritter, F. E., Haynes, S. R., & Cohen, M. A. (2004). CaDaDis: A tool for displaying the behavior of cognitive models and agents. In *Proceedings of the 13th Conference on Behavior Representation in Modeling and Simulation*. 04-BRIMS-032. 192-200. Orlando, FL: U. of Central Florida. http://www.sisostds .org/cgf-br/04Brims [link checked 4 June 2004].

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, UK: Cambridge University Press.

van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The Think Aloud Method: A practical guide to modelling cognitive processes*. London and San Diego: Academic Press.

Wickens, C. D., Gordon, S. E., & Liu, Y. (1998). *An introduction to human factors engineering* (3rd ed.). Addison-Wesley: New York, NY.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Prentice-Hall: Upper Saddle River, NJ.

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: Freeman.

Wiesmeyer, M. D. (1992). *An operator-based model of human covert visual attention*. Ph.D. thesis, U. of Michigan.

Yen, J., Yin, J., Ioerger, T. R., Miller, M. S., Xu, D., & Volz, R. A. (2001). CAST: Collaborative agents for simulating teamwork. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01)*. 1135-1142. Los Altos, CA: Morgan Kaufmann.

Yost, G. R., & Newell, A. (1989). A problem space approach to expert system specification. In *Eleventh International Joint Conference on Artificial Intelligence*. 621-627.

Zachary, W., Jones, R. M., & Taylor, G. (2002). How to communicate to users what is inside a cognitive model. In *Proceedings of the 11th Computer Generated Forces Conference.* 375-382, 02-CGF-114. Orlando, FL: U. of Central Florida.

Also see http://acs.ist.psu.edu/papers/ for online copies of papers by Ritter.