A Method for Building Models of Expert Cognition in Naturalistic Environments

W. Korey MacDougall (KoreyMacdougall@gmail.com)
Matthew Martin (MattMartin256@gmail.com)
Nathan Nagy (n.c.nagy1@gmail.com)
Robert L. West (rlWest@gmail.com)
Institute of Cognitive Science, Carleton University, Canada
Ottawa, ON K1S5B6 Canada

Abstract

We discuss a method for creating models of expert cognition and behavior in naturalistic environments. The method consists of video annotation and iterative model tracing, as well as a commitment to making all data and components of the model available for independent validation.

Keywords: Cognitive modeling; macro-cognition; expertise; expert cognition; methodology.

Introduction

Cognitive modeling draws largely upon theories and methods from psychology and artificial intelligence. While the insights and approaches developed in these fields are leveraged to great positive effect by cognitive modelers, there are limitations that arise from an overly strict adherence to the practices of these sister fields. Among these is the difficulty of modeling behavior and cognition in naturalistic environments by using laboratory-based methods. The reasons for this issue, and the approach we are using to address it are discussed below.

Laboratory Methods and Macro-cognition

Studying behavior and cognition in real-life scenarios is difficult. Environments can be chaotic, behavior can be inconsistent, and there can be an overwhelming number of variables involved. Laboratory experimentation is largely aimed at limiting this complexity and helping researchers to isolate and examine more precisely the factors they are interested in. This is achieved through simplification of the task environment, repetition of tasks, and the averaging of data across trials and participants.

This methodology has proven powerful and generative, but questions have been raised about the difficulties of “scaling up” findings from laboratory experimentation to explain cognition as it occurs in the real world (Klein et al., 2003). These limitations are particularly relevant in the study of expert behavior and cognition. In natural environments, expert behavior is dynamic, adaptive, and often idiosyncratic. These are important elements that are often obscured by traditional laboratory paradigms.

As an example, consider the difference between the study of chess players as it traditionally occurs in laboratory paradigms and the observation of a chess master playing a game at home. In lab studies, players may be presented with chess positions and asked to recall as many pieces as possible, after which reaction times and error rates can be measured. Alternatively, they may be shown a position and asked to make a single move, then asked how many candidate moves were considered, or to speak aloud while deciding what move to play. This is a form of protocol analysis commonly used to supplement experimental work (Ericsson, 2006).

Using the above methods, the motivation is often to examine how many elements can be “chunked” into a single memory representation, or to examine how deeply or broadly a player searches in choosing their next move. This is a reductive approach that splits the task performance into pieces in order to isolate and better understand those pieces, with the (possible) intention of later combining the components into a more holistic picture of the underlying cognition. This subsequent recombination can be directed in two ways: either to form a more complete picture of chess cognition *per se*, or to inform a broader theory of cognitive functions, such as memory encoding or pattern recognition, which in the general case are not specifically tethered to chess. Part of our motivation in creating the method described here is skepticism about whether this process of division and subsequent recombination can lead to an understanding of the cognitive system as a whole, particularly as it applies to understanding situated, real world, expert cognition.

The tension between the power of laboratory methods and the complexity of situated expertise has been addressed in a number of ways (Kieras & Meyer, 2000; Klein et al., 2003; Williams, 2006; West & Nagy, 2007). One approach that is useful in the context of cognitive modeling is to distinguish between micro-cognition and macro-cognition. Micro-cognition refers to those mental operations that are typically studied in cognitive psychology experiments and which are thought to be invariant and underlie all of cognition (Klein et al., 2003). These include such functions as memory encoding and retrieval, and serial versus parallel attentional mechanisms (Klein et al, 2003). Macro-cognition, on the other hand, refers to cognition as it occurs in naturalistic environments, and includes such high-level operations as complex decision making, resource allocation, team co-ordination, and responding to non-routine circumstances (Schraagen, Militello, Ormerod, & Lipshitz, 2008; West & Nagy, 2007).

The method we are using is aimed at elucidating macro-cognitive processes and therefore eschews the siloing of the component functions, as is common in studies of micro cognition. Instead, we observe the expert performing the task as they naturally would, with minimal coaching or restrictions, and attempt to identify the components of the task afterward, in collaboration with the experts. We wish to point out that we are not creating this as an *alternative* to lab based micro-cognition research. Instead, our approach is intended for use *in conjunction with* traditional micro-cognitive research methods.

Method

Motivation, Philosophy and Scope

There are two guiding motivations for this methodology. The first is the belief in the value of unification in modeling, and the second is what we perceive as the importance of integrating and making explicit the relations between experimental design, data analysis, and theory construction. Concerning the first motivation, we agree with Newell's (1973) argument that, if we are ever to understand cognitive systems, the research community must attempt to integrate its efforts and avoid an unbounded proliferation of unconnected models.

As for the second point, we want to encourage explicit attention to the relations between theory and method in cognitive modeling and experimental work. In particular, for this discussion it is important to distinguish between (1) systems for building computational models (e.g., ACT-R, GOMS, SOAR); (2) methods for creating models using these systems (e.g., task analysis, cognitive walkthroughs, ethnology); and (3) methods for evaluating the resulting models (e.g., hypothesis testing, model fitting). What we are proposing is a method for *evaluating* models. It can be used with any computational modeling system and any methodology for generating models within these systems.

The intended scope of this methodology is the study of real world expertise, specifically in those domains for which some, but not all, behavior and cognition is routine. In the case of novices, behavior is generally too variable to be modeled using this approach. As practitioners develop proficiency in their domain, they generally converge upon optimal solutions (Shanteau & Hall, 2001), and thus we observe more consistency at higher levels of expertise than at lower levels (note: this methodology is not intended for "creative" expert domains such as music composition or fiction writing, where no convergence on an optimal process is expected). Due to factors such as chaotic environments, individual differences, the actions of co-workers, unexpected events, and/or the need to multi-task, experts in the same field do not always behave in the same way. We are interested in the middle ground between behavior that is fully routine and repetitive on the one hand, and that which seems entirely unsystematic on the other. We argue that this is the zone in which most real-world experts operate.

Our methodology is more akin to systems engineering practice than it is to experimentation. We are not attempting to generate and test hypotheses as we would in lab-based experiments. Rather, we are attempting to develop and refine models until they adequately capture the range of relevant behaviors and cognitive operations. This approach is more consistent with a Lakatosian scientific framework (Lakatos, 1970) than with a Popperian one (Popper, 1963). In short, the process of evaluating our models rests upon iteration rather than falsification.

While we respect the importance of falsifiability in theorizing, we must also be clear about when it is appropriate or possible. For example, it is problematic to use falsification to evaluate the validity of cognitive architectures, such as ACT-R, GOMS, or SOAR (Cooper, 2007; Newell, 1973). Although some models built in these systems can be falsified, the architectures generally cannot be falsified because there are usually multiple ways to model the same task within a single architecture. In other words, the model can be adjusted to fit the data.

Likewise, we argue that falsification is problematic for evaluating models of real world expert behaviour, but for different reasons. Specifically, although we are concerned with evaluating specific models, and not the architectures in which they are built, the naturalistic behaviour of experts across time is different each time they are observed, even for the same individual. Of course, many of the component behaviours, or unit tasks, are the same from scenario to scenario, and these can be isolated and studied in the lab, but this is not what we are evaluating. Our interest is in evaluating whether a model can realistically account for the sequence of decisions and behaviours as each different scenario unfolds. Lab-based hypothesis testing is inappropriate here because it is based on averaging across the same sequence of behaviours repeated within and/or across individuals, with no variations in the environment.

Lakatos (1970) defines a program of research as scientific if it is making progress over time, where progress may be demonstrated in multiple ways. For example, progress may include the discovery of new phenomena, falsification, hypothesis confirmation, increased parsimony, theory unification, counterfactual predictions, etc. Our method is based on two criteria for progress: (1) an increase in the amount of data accounted for by the model, and an increase in the percentage of times the model correctly predicts the next action of the human expert, and (2) an increase in the scope of the model, i.e., as we collect more and more samples of expert behaviours, the same model must cover all of them without any parameter changes.

Overview

The method we have been developing is an iterative, collaborative approach to creating macro-cognitive models. The process involves recording video footage of experts performing in naturalistic environments then using this information to construct a model of the task. Using this model as a base, we use an iterative model-tracing procedure to improve it until it is sufficiently robust to predict all or most of the high-level behavior observed. We do this using freely available tools and make our data and models available to other interested parties. The process is laid out in more detail below.

Procedure

1 - Video capture The first step is to collect video footage of experts performing in a naturalistic environment, as well as documentation about the task and interviews with experts. Rather than constructing a simplified task environment and attempting to isolate components of the task performance, we aim to have the experts demonstrate their skills in the messiness and complexity to which they are accustomed and which forms the necessary background against which their training and expertise are normally expressed. This step is similar to techniques used in cognitive task analysis (Kieras & Meyer, 2000) and cognitive ethnography (Williams, 2006).

2 - Task Model Construction Once video footage has been collected, we review it with the experts and attempt to determine patterns and regularities in task performance. We ask the experts to tell us what their goals and sub-goals were at each given point, what their strategies for accomplishing these goals were, and to identify which elements of the environment were relevant in their decision making (note that this process can be begun before video data is collected).

3 - Cognitive Model Construction Once we have created a task model, we construct two separate but inter-related models: a cognitive process model capable of completing the tasks, and a perceptual model, which we have termed the situational awareness (SA) model, that describes what the agent pays attention to in the environment and how these environmental cues are combined into a meaningful interpretation. These two models are linked in that the process model relies on the SA model for a meaningful interpretation of the environment and the SA model relies on the process model to provide context (e.g., in terms of the current goals of the agent), which is used to interpret raw environmental cues to create situational awareness.

The framework we are using to inform this step is called Sociotechnical GOMS, or SGOMS (MacDougall, West, & Hancock, 2012; West & Nagy, 2007), which is an extension of the GOMS modeling framework (John & Kieras, 1996). However, our method is not necessarily tied to any particular theory of cognition and therefore we will not discuss SGOMS in detail.

4 – Video Annotation and Model Tracing Once we have constructed the two models, we use them to annotate the video footage we have collected. We identify which actions are being undertaken at each point in the video, what the current goals and constraints driving behavior are, and which elements of the context are relevant in decision making. As these are determined, we note on the video which actions are being undertaken and specify their time course. To create these annotations we are using the ANVIL Video Annotation software (Kipp, 2010). See Figure 1 in for a screenshot of video that has been annotated in ANVIL.

Figure 1: ANVIL-annotated video frame of gameplay.

The annotation procedure is accomplished through an iterative process of model tracing. To do this, we first annotate the video by noting when behavioural elements related to the cognitive model appear on the video. Then we annotate the video with regard to the SA model. This process occurs in multiple cycles or iterations. We repeatedly make additions and deletions to the models in order to more accurately capture the range of behaviors and relevant contextual elements demonstrated in the dataset.

In the refining process, we use an adapted form of model tracing, a practice that has been used to positive effect by designers of intelligent tutoring systems (Koedinger & Anderson, 1997; VanLehn, Freedman, & Jordan, 2000). In effect, we try to determine at each point in the task performance whether the observed human behavior is consistent with predictions made by the model. We do this by assessing whether the model could have *reasonably* (see the discussion section for more on this term) chosen the same action as the human agent did, given the states of both the SA and cognitive models' allowable responses to that state. If the models cannot account for the observed behavior, we attempt to modify them. This is one of the ways in which this method differs significantly from hypothesis-based experimentation: we do not conduct the annotation process in order to test whether the current build of our model is correct or incorrect, but rather with the intention of improving the existing model so that it more accurately and parsimoniously predicts the observed behavior. In a sense then, the model informing the annotation can be viewed as a "rolling hypothesis" which is updated and refined with each iteration over the video.

When making modifications to the model, we create a second "branch" of the model (as in software engineering), and test the new configuration against the previous one. If the new additions or deletions improve the accuracy of the model, they are maintained, otherwise they are rejected and the initial model is retained. We determine whether an iteration is an improvement upon the previous model by noting the number of times each version fails to predict what the expert did. We consider the model construction process complete when further iterations cease improving the accuracy of the model. Some potential difficulties with this part of the method, such as the risk of over-fitting the model or of having an unbounded number of possible actions within the model, are examined in the discussion section.

5 – Model and Data Release Once we have finished developing a model and have used it to annotate video footage, we release online both the model and the annotated data (video footage) to other interested parties. This is, we think, a crucial component of cognitive modeling at the communal level. It encourages transparency and allows for more rigorous peer evaluation of research claims, and it facilitates collaboration between investigators. It also encourages data and model re-use.

None of the various elements that we have combined are new. Our contribution lies, we hope, in demonstrating the scientific potential of embedding iterative model building in a systematic, explicit methodology for evaluating models of real world expertise.

Example: Video Game Playing

We have used the method described here to construct models of video game playing, professional mediation, chess playing, and professional cooking. We will describe one of these cases, namely a model of playing Gears of War 3 (Activision), a third-person shooter game for the Microsoft XBOX 360.

In order to construct the model, we had several expert players play the game while we recorded the screen. Afterwards, we asked the individuals to discuss their strategies and thought processes while playing, and began to construct the task model. Once we had an idea of what they were paying attention to in the environment, and how they were making decisions, we began constructing the SA model and the cognitive model.

Figure 2, below, represents the process schematically. On the top is the video frame from Figure 1; this depicts what players would see on screen, and was the video data that formed the basis for the annotations. On the bottom left is a representation of the cognitive model. This contains cognitive and behavioral actions, such as “find cover”, “engage enemy”, or “assess threat” along with the conditions under which they can occur. The SGOMS model also covers high level planning and dealing with unexpected interruptions. The visualization of the cognitive model depicted in Figure 2 is output from software that we have developed in-house for visualizing these models; the software can be downloaded at https://github.com/mattmartin256/SGOMS_GUI. On the bottom right is a representation of the SA model, which lists the important elements of the environment that are attended to. Examples of these elements include the number of enemies on the screen, whether ammunition is running low, and the state of the character’s health. The blue arrows between components represent the fact that the construction process is iterative and that each component is used to modify and refine the others.

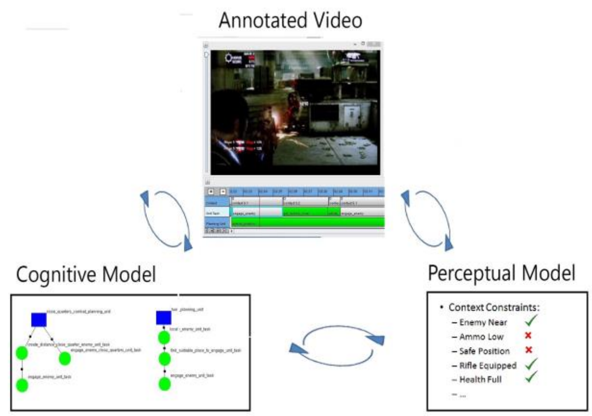

Figure 2: Schematic of Model Construction Process.

Discussion

The methodology presented here is a work in progress. We are attempting to develop a systematic way of modeling expert behavior in complex, real-life scenarios. Such a methodology would be valuable, we argue, both for basic cognition research and to inform the design of socio-technical systems. There are a number of potential difficulties, however, with such a method and we anticipate a number of criticisms here.

The principle difficulty, and the one which this methodology is most explicitly attempting to address, is that there can be an overwhelming amount of complexity in naturalistic environments, and determining what is important or relevant is not straightforward. Laboratory methods are aimed at carving out a tractable section of cognition or behavior, to save the investigators from being forced to address everything all at once: some sub-set of phenomena is selected as important. Determining what is to be included in a cognitive model or experimental design is not atheoretical or pre-theoretical though. Each experimental design or modeling framework necessarily includes certain elements and excludes others. Any chosen methodology thus “smuggles in” theoretical assumptions about what ought to be paid attention to and what we can safely ignore. This is the problem of deciding on one’s “unit of analysis”, which is a common problem in all scientific disciplines (Hutchins, 2010).

The various approaches for studying expertise define the unit of analysis in many different ways according to the methods and theories being employed. These may include error rates and reaction times (Burkhardt, Détienne, & Wiedenbeck, 1997), memory recall tasks (Vicente & Wang, 1998), eye movement patterns (Reingold, Charness, Pomplun, & Stampe, 2001), and verbal protocols (Greenwood & King, 1995), among others. The unit of analysis we wish to use is the interaction of an expert (or group thereof) with a complex socio-technical system. We are trying to accommodate this complex unit by bridging the methodologies of experimental psychology approaches that use rich environmental and behavioral descriptions, such as cognitive ecology and anthropology (Bender, Hutchins, & Medin, 2010; D'Andrade, 1995). In essence, we are trying to combine the rigor and predictiveness of process modeling with the richness of ecological studies.

The second difficulty is that individuals are often unable to vocalize what they know (Clark, Yates, Early, & Merriënboer, 2008). We accept that much of the cognitive activity occurring "under the hood" will be invisible and may be unavailable for reporting by the expert. We thus do not assume that the input from our participant experts is the final word on what they are doing mentally. At the same time, however, we believe that this feedback from experts is a desirable component in modeling expertise, and a useful starting point for developing models of expert cognition. Also, it is important to note that the modeling approach used will affect the interactions with the experts. For example, we used SGOMS so our interactions with the experts were naturally geared toward eliciting the information and structures needed to build this type of model.

A third challenge with this method is the degree to which human judgment is required in the construction and evaluation of these models. When deciding whether to add or remove an element from a model, or in judging whether the model has accurately predicted a sequence of behavior, we must rely on the modellers' knowledge and discrimination, and these cannot be perfectly formalized. In other words, the evaluators must decide whether the model could have "reasonably" predicted each decision and action, given the elements in the cognitive and perceptual models. Because we are specifically interested in complex environments in which there is significant behavioral variability between participants and trials, we must use judgment in determining whether two instances of action are equivalent according the model. For example: in the game play scenario presented above, no two instances of the action "take cover" will be exactly the same on the screen, so we must be capable of abstracting from the data to equate the two instances.

Our stance on this issue is informed by Herb Simon's (1969) "ant on the beach metaphor". This states that the observed behavioral complexity and variability of an agent is often the result of the environment in which the agent acts, and does not originate within the agent itself. In the case of the ant, the insect is a rudimentary cognitive-behavioral system. Watching an ant navigate a sandy beach, it may seem that the ant is moving in complex patterns, when, in reality, it may only be obeying the simple heuristic of "do not climb hills". The point that we take from this is that superficially distinct behaviors may reflect the same underlying cognitive processes. It is in determining whether such equivalence exists between instances of behavior that the role of judgment comes into play in this methodology. Here, the public availability of the cognitive and perceptual models along with the annotated videos plays a crucial role. The claim that a judgment was reasonable must stand up to public scrutiny.

Another consideration in the use of this methodology is the difficulty of choosing which elements to include in a given model. We need to negotiate between two extremes: over-fitting and unbounded growth. In the former case, we want models to be capable of accurately predicting behavior by the collection of experts studied, and a model tied too specifically to a single instance or agent will fail to meet this goal. In the latter case, we want to avoid the temptation of endlessly adding elements to the model whenever something unexpected occurs. This is connected to Simon's ant metaphor: we need to determine when superficially dissimilar behaviors represent the same underlying mechanism, because without such abstraction and equating, the models will quickly become bloated and unwieldy. The final goal is to develop models that are informative, predictive, and lean, and this requires a balance between specificity and generality.

One way of evaluating progress on this path arises from our goal to create one unified model that applies across all instances of a field of expertise. Following from our use of branching and the comparison of new models with old, we would expect our current model to be backwards compatible across all the videos used up to that point. If the model has been over-fit it will not show a consistent advantage across all of the videos. With each iteration the demands on the model are actually increased. Eventually, diminishing returns on adjusting the model would signal that it is about as good as it will get. At this point the model and the annotated videos can be used as a benchmark to evaluate alternative models against.

In terms of quantitative evaluation, we are currently using as our metric the percentage of correct predictions the model makes of the expert's next action (this measure was also used in West & Nagy, 2007). However, we are working on other, more detailed ways of quantifying how good the models are. For example, in some cases the model makes a single prediction of what is likely to come come next, but in other cases the model allows for more than one possible next action. Currently, we are experimenting with incorporating the number of predicted next actions at each point into our quantitative measure. For example: if two models are equally predictive in terms of the percentage of actions accurately predicted, but one model regularly predicts a greater number of next possible actions, that model ought to score lower, as it is a less lean (more bloated) model of the expert behavior.

Conclusion

We have presented a methodology for modeling expert behavior and cognition in complex, naturalistic environments. Such a technique will, we hope, be valuable in furthering our understanding of expertise and situated cognition, and may also be useful in improving the design of socio-technical systems, such as emergency operations centers or aircraft cockpits. We support an open-source approach to scientific research, and hope that explicit attention to methodology, along with the open sharing of tools, data, and models will facilitate collaboration among researchers and the development of more unified, comprehensive cognitive models.

References

Bender, A., Hutchins, E., & Medin, D. (2010). Anthropology in cognitive science. *Topics in Cognitive Science*, *2*, 374–385.

Burkhardt, J. M., Détienne, F., & Wiedenbeck, S. (1997, January). Mental representations constructed by experts and novices in object-oriented program comprehension. In *Human-Computer Interaction INTERACT'97*. Springer US.

Clark, R. E., Feldon, D., van Merriënboer, J. J., Yates, K., & Early, S. (2008). Cognitive task analysis. *Handbook of research on educational communications and technology, 3*, 577-593.

Cooper, R. P. (2007). The role of falsification in the development of cognitive architectures: insights from a Lakatosian analysis. *Cognitive Science*, *31*, 509–533.

d'Andrade, R. G. (1995). *The development of cognitive anthropology*. Cambridge University Press.

Ericsson, A. K. (2006). Protocol Analysis and Expert Thought: Concurrent Verbalizations of Thinking during Experts Performance on Representative Tasks. In *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge University Press.

Greenwood, J., & King, M. (1995). Some surprising similarities in the clinical reasoning of "expert" and "novice" orthopaedic nurses: report of a study using verbal protocols and protocol analyses. *Journal of Advanced Nursing*, *22*, 907–913.

Hutchins, E. (2010). Cognitive ecology. *Topics in Cognitive Science*, *2*(4), 705–715.

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

Kieras, D. E., & Meyer, D. E. (2000). The role of cognitive task analysis in the application of predictive models of human performance. In *Cognitive Task Analysis*. Psychology Press.

Kipp, M. (2010). Anvil: The video annotation research tool. In *Handbook of Corpus Phonology*. Oxford University Press.

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *Intelligent Systems, IEEE*, *18*(3), 81-85.

Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*, 30–43.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In *Criticism and the Growth of Knowledge*, *4*, 91–196

MacDougall, W., West, R., & Hancock, E. (2012). Modeling Multi-Agent Chaos: Killing Aliens and Managing Difficult People. *Proceedings of the Thirty-sixth Annual Conference of the Cognitive Science Society* (pp. 2603-2608). Hillsdale, NJ: Lawrence Erlbaum Associates.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In *Visual Information Processing*. Academic Press.

Popper, K. R. (1963). Science as falsification. In T. Schtick (Ed.), *Conjectures and refutations*. Mountain View, CA: Routledge and Keagan Paul.

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: evidence from eye movements. *Psychological Science: A Journal of the American Psychological Society, 12*, 48–55.

Schraagen, J. M., Ormerod, T., Militello, L., & Lipshitz, R. (Eds.). (2012). *Naturalistic decision making and macrocognition*. Ashgate Publishing Ltd.

Shanteau, J., & Hall, B. (2001). What does it mean when experts disagree. In E. Salas, & G. Klein (Eds.), *Linking expertise and naturalistic decision making*. Mahwah, N. J.: Lawrence Erlbaum Associates.

Simon, H. A. (1996). *The sciences of the artificial* (Vol. 136). MIT press.

VanLehn, K., Freedman, R., Jordan, P., Murray, C., Rosé, C. P., Schulze, K., Shelby, R., Treacy, D., Weinstein, A., & Wintersgill, M. (2000). Fading and deepening: The next steps for Andes and other model-tracing tutors. In Gauthier, Frasson, VanLehn (Eds.), *Intelligent Tutoring Systems: 5th International Conference: Vol. 1839. Lecture Notes in Computer Science*. Springer-Verlag Berlin & Heidelberg GmbH & Co. K.

Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review*, *105*, 33–57.

West, R. L., & Nagy, G. (2007). Using GOMS for Modeling Routine Tasks Within Complex Sociotechnical Systems: Connecting Macrocognitive Models to Microcognition. *Journal of Cognitive Engineering and Decision Making*, *1*(2), 186–211.

Williams, R. (2006). Using cognitive ethnography to study instruction. *Proceedings of the 7th International Conference on Learning Sciences*, 838–844. International Society of the Learning Sciences