Controlled Exploration of Alternative Mechanisms in Cognitive Modeling

Rita Kovordányi (ritko@ida.liu.se)
Department of Computer and Information Science
Linköpings Universitet, SE-581 83 Linköping, Sweden

Abstract

Overt cognitive behavior arises through a complex interaction between internal, not directly observable, cognitive mechanisms. As there may be several ways of achieving the same overt behavior, it is intrinsically difficult to find the “correct” model. One way to proceed however is to uncover the causal dependencies between a particular configuration of cognitive mechanisms and simulated overt behavior. This can be achieved in controlled simulation experiments where every combination of potentially important cognitive mechanisms is systematically tried out. To illustrate this point, we briefly describe an application of the two-level factorial simulation design on a modeling project in mental imagery. We conclude by discussing the potential of the method as a tool for reliable incremental model development.

Introduction

The general objective of modeling and simulation is often to correctly predict real-world system performance. In addition to this, cognitive modeling aims at discovering the true nature of cognition (Kieras, 1987; Anderson, 1993; Newell, 1990; Kosslyn, 1980, 1994; Kosslyn et al., 1979). Ideally, this would presuppose either that a cognitive model can be rejected as invalid with respect to empirical data, or that a cognitive model, or a particular cognitive mechanism, can be singled out as being valid within a given theoretical setting.

However, behavioral data often do not cover every necessary aspect of a cognitive phenomenon or are qualitative in nature, and may thus be consistent with a range of possible accounts. This open-endedness poses a severe problem in cognitive theory construction and model building, a problem which is commonly known as the identifiability problem: “The thorny issue of how we can know [that we have arrived at] the correct theory” (Anderson, 1993, p. 10).

Several ways of dealing with this problem have emerged during decades of modeling practice. First, the empirical basis for model construction can be broadened to increase the number of constraints and thereby pin down the gross structure of possible cognitive models. Within the space of possible models which is left, often ad hoc or heuristic search is employed to find a model which satisfies the full range of data (Kieras, 1985). In general, this method increases the probability that the model found is also “correct” in a broader sense.

Second, unified architectures of cognition are incrementally constructed in a team effort and evolve through years of development to accumulate a wide range of empirical data. These architectures outline the main processing subsystems and the flow of processing in the cognitive system, and in this way support the development of specific, lower-level models (Rosenbloom et al., 1993; Anderson, 1993).

Overt cognitive behavior arises from a complex interaction between internal cognitive mechanisms. Even when model development is guided by assumptions about the overall cognitive architecture, it may be difficult to pin-point which of several possible mechanisms is responsible for a set of empirical observations. For example, should the empirically observed reaction time and error-rate effects of “attending” to visual stimuli be attributed to early or late selection in the visual system, assuming that visual perception is implemented in a hierarchy of mutually interacting stages of processing?

In general, there is a need to untangle the complex interaction between hypothetical cognitive mechanisms. On the one hand, one would like to establish a causal link between central mechanisms and their contribution to overall model behavior. On the other hand, one would like to identify those mechanisms, which either give rise to invalid behavior, or do not significantly contribute to overall model performance. Strictly speaking, this entails an experimentation with cognitive models using an experimental design where every cognitive mechanism in the chain or network of mechanisms involved in a cognitive task is systematically varied so that alternative implementations of individual mechanisms can be fully cross-combined.

For practical reasons, high-dimensional experimental designs are avoided in real-world, psychological experiments. However, in general such practical limitations do not apply to a computer simulation environment. Yet, the full factorial design (cf. section on ‘The two-level factorial design’ below) is not employed in cognitive modeling.

In a modeling project on mental imagery (Kovordányi, 1999b), we have adopted this approach and have systematically simulated alternative embodiments of a generic interactive activation model (McClelland, 1979; McClelland and Rumelhart, 1981, 1994/1988; Rumelhart and McClelland, 1982). Based on our experience with this project, we would like to point to the potentials of this method.

Simulating cognitive models in a controlled experimental setting

The advocated method for exploring cognitive models may be conceived of as the equivalent of running a high-dimensional real-world experimental design with a multi-way analysis of co-variation (multi-way ANOVA). In this sense, the space of cognitive models is used as a virtual environment for experimentation: The structure of this environment is partially fixed by what we call the model framework. “The independent variables” correspond to those aspects of the cognitive model which cannot be specified in advance, but which may be potential determinants for the model’s overall behavior. “The dependent variable” constitutes a measure of model performance which, for purposes of model validation, should correspond to experimentally observed behavior in human subjects. Experimentation through systematic model simulation aims to shed light on how some of the “a priori” unknown aspects of the partially specified model interact in affecting the model’s behavior, and most importantly, whether a specific combination of model properties produces valid model behavior.

The two-level factorial design

Systematic exploration of alternative model instances can be organized according to a full two-level factorial design (Law and Kelton, 1991; Box et al., 1978). This design emphasizes that the question of which model parameters are causally involved in a particular type of simulated behavior can be answered only if all parameters have been fully cross-combined. In order to keep down the computational cost of exploring all parameters, parameter values are varied between a predetermined min- and max-value, in what is called a two-level factorial design.

Note that, for the above reasons, if some model parameters were to be fixed at a given “reasonable value” in order to keep down simulation complexity, the power of the simulation design would diminish. Strictly speaking, such simulations cannot validate conclusions about which model properties are causally involved in the simulated behavior. Simply expressed, parameters may have been fixed at a value where they in fact interact with the central parameters of the model. Hence, for example, if no effect is obtained when the value of one of the central parameters is varied, this could in fact hide a significant negative effect, which is positively modulated by a peripheral parameter, which has been fixed.

Ideally, for a problem with k degrees of freedom, the minimal number of simulations which needs to be run in order to detect causal dependencies between model parameters is 2^k. However, if the number of simulations turn out to be unmanageably large, a fractal two-level factorial design may be the used instead of a full design (cf. Law and Kelton, 1991; Box et al., 1978). In these designs, peripheral parameters are not fixed at an ad hoc value, but are instead defined dynamically to be a function of other, more central parameters.

In addition to providing a minimally sufficient basis for detecting causal relationships in the simulation results, using a two-level factorial design renders the analysis of simulation results computationally simple. A simulation where k parameters are varied is captured in a design matrix of size 2^k x k containing +s and –s representing low and high parameter values (cf. Law and Kelton, 1991; Box et al., 1978). The way the matrix is set up, each row will represent a unique combination of parameter values, which in turn corresponds to a particular simulation run. As the design matrix is regular, it is easy to set up. In addition, once it is computed, the same matrix can be used to control the simulations and to conduct data analysis.

To illustrate the latter case, if the possible interaction between parameters p_1, p_3, and p_7 are inquired, columns 1, 3, and 7 of the design matrix are multiplied value-by-value, and then multiplied with the set of simulation data. The effect of these multiplications is that the correct signs will be added to the data column. A final summation of all the signed entries in the data column, divided by 2^{k-1}, where k denotes the number of model parameters, yields the desired mean interaction of the parameters involved (cf. figure 1).

run	par 1	par 2	par 3	sim. result
1	–	–	–	R_1
2	–	–	+	R_2
3	–	+	–	R_3
4	–	+	+	R_4
5	+	–	–	R_5
6	+	–	+	R_6
7	+	+	–	R_7
8	+	+	+	R_8

Figure 1: Example of a two-level full factorial simulation design matrix for three parameters. Each row in the matrix denotes a unique combination of parameter values. The last column in the design matrix designates the outcome of simulating a model (instance) for that particular parameter combination.

Our modeling project

In our investigation of mental imagery, a full two-level factorial design was used where all parameters not inherently dependent on each other were cross-combined (Kovordányi, 1999b, 2000). While variations in the effect of several possible factors, such as the effect of mental image fading, were taken into account, simulation data analysis was centered around uncovering the effect of focusing early versus late selective attention on part of a mental image in a mental image reinterpretation task. As the empirical results of Finke and colleagues (Finke et al., 1989) and Peterson and colleagues (Peterson et al., 1992) used for model validation were qualitative, no attempt was made to optimize the models towards these data (Kovordányi, 2000). Model validity was instead defined qualitatively, and served as a means for “filtering out” invalid model instances.

Parameterization of the model design space

The interactive activation model used in our project (cf. Kovordányi, 1998, 1999a) drew its main architectural components from the comprehensive model of mental imagery forwarded by Kosslyn (1994; Kosslyn et al., 1979; Kosslyn et al., 1990). This model framework enabled us to capture all basic assumptions made at a higher, theoretical level, while enabling a systematic search for algorithmic details, which were left open by the theoretical and empirical basis.

How should an underconstrained model be partially specified so that it allows for a natural variation of model properties? One approach, used in our modeling project, is to set up a generic model framework as a localist network, and let each node in this network encode a holistic property or feature of the modeled phenomenon. In the case of visual perception, one kind of holistic property would be, for example, the individual line segments, which make up more complex line drawings.

One example of localist networks is the interactive activation model developed by McClelland and Rumelhart (McClelland, 1979; McClelland and Rumelhart, 1981, 1994/1988; Rumelhart and McClelland, 1982). In these models, the localist nodes are arranged into reciprocally connected layers of processing, thereby further increasing the structure and penetrability of the model. Units within the same processing layer are assumed to have the same inhibitory/excitatory connection weights. In such a model framework, model parameters can be naturally expressed as connection weights, activation thresholds, resting levels, or simply as "control flags". These flags could, for example, control whether an individual simulation run should be initiated top-down or bottom-up in the interactive network.

Model parameters can arise naturally also in symbolic models. Parameters in these models could be represented as alternative (sets of) production rules, or simply alternative definitions (fnc1 – fnc2) of a cognitive mechanism together with some means for activating them at run-time. Hence, in essence, any modularly built computational model can be parameterized with a minimal overhead cost.

Simulations

Our model framework for mental imagery encompasses three mutually interacting layers of processing (figure 2). At the lowest level, the visual buffer contains detectors for oriented line segments. At the next stage, these feature detectors can evoke (and get feedback from) simple geometric patterns, such as composite lines or triangles, which are stored in visual long-term memory. At the highest level of processing, geometric patterns are combined into abstract concepts stored in amodal, associative long-term memory. In addition to between-layer connections, there is lateral, within-processing-level inhibition between mutually inconsistent (groups of) units. Interpretation in this system entails the dynamic establishment of a correspondence between low-level and higher-level representations.

We simulated mental and perceptual reinterpretation of two composite line drawings from Finke and colleagues (1989, exp. 1). Possible interpretations of these figures were limited to a small set of predefined geometric forms and abstract concepts. For example, possible interpretations of the first figure, formed from an upper case 'H' superim-

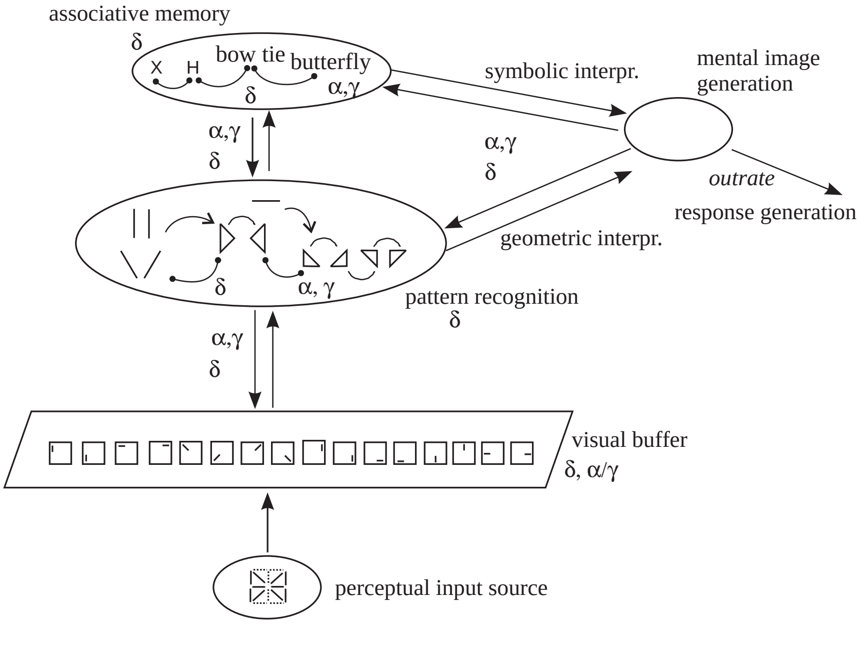


Figure 2: Communication and control structure of our model. Model parameters are shown as tags attached to the corresponding connection or subsystem. Note that model performance is expected to depend not only on how parameters are set, but also on whether the system is initiated top-down or bottom-up. These the two ways of initiating the system correspond to mental imagery and visual perception, respectively.

posed on an upper case ‘X’, were limited to “four small equilateral triangles”, “two large isosceles triangles”, “a butterfly”, “a tilted hourglass” and “a bow-tie”.

As layers in the system were reciprocally interconnected, simulations could be initiated either top-down or bottom-up. This made it possible to compare reinterpretation performance in visual perception and in mental imagery. When simulations were run in mental mode, a chosen symbolic concept was activated in associative long-term memory, and this activation was projected into the visual buffer, where an activation pattern emerged which represented a visual mental image. When simulation was run in perceptual mode, visual input entered the system at the visual buffer, and was forwarded through consecutive stages of processing, and matched to geometric patterns and abstract concepts. One of these patterns or concepts was selected for verbal report.

Simulations were run through four phases: Mental image generation, followed by mental image reinterpretation, continued with a corresponding perceptual image build-up of the same line-figure, followed by perceptually based reinterpretation. Each simulation was run for 10 simulated seconds, in discrete simulation steps of 50 ms.

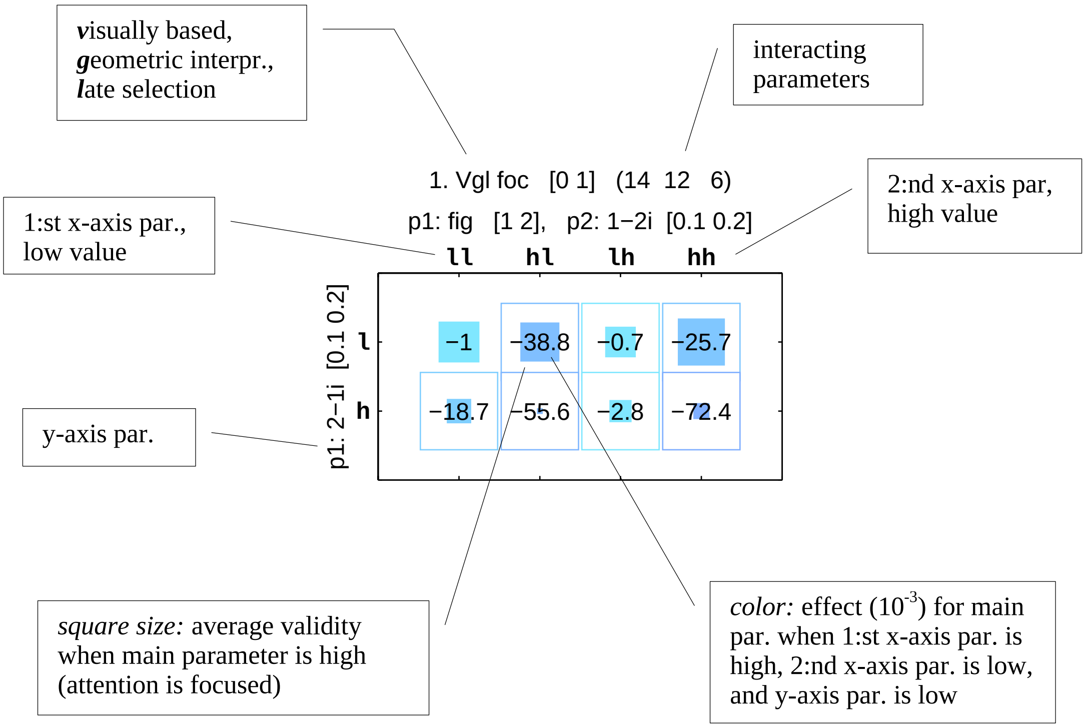
Two configurations of the model framework were scrutinized: One where attentional selection occurred late, affecting processing at the level of associative long-term memory, and one where selection occurred early and directly affected the contents of the visual buffer. For these model configurations, the effect of focusing attention (versus not focusing attention) was investigated, taking into account the interaction effects that arose between this central, and other peripheral model parameters.

Data analysis

In our project, data analysis was based on semi-automatic preparation of the raw simulation data. The prepared data were then visualized. The aim was to facilitate the discovery of significant parameter interactions, and in addition provide a basis for estimating model validity for the different parameter combinations. Below we briefly describe the key stages of this process.

Identification of interacting parameters

Activation levels of all response units in the interactive activation network were measured for each simulation run, that is for each parameter combination (cf. Kovordányi, 1999b). From these activation values the corresponding probability for mental reinterpretation was calculated. Mental reinterpretation rates were considered valid if they qualitatively matched the reinterpretation rates obtained by Finke and



***v**isually based, **g**eometric interpr., **l**ate selection*

interacting parameters

2:nd x-axis par, high value

1:st x-axis par., low value

y-axis par.

square size: average validity when main parameter is high (attention is focused)

color: effect (10^{-3}) for main par. when 1:st x-axis par. is high, 2:nd x-axis par. is low, and y-axis par. is low

Figure 3: Example visualization of the simulation data. Data values are color-coded to support the understanding of interaction patterns. The area of the square markers reflects the validity of the underlying parameter combinations.

colleagues (1989, exp. 1), and Peterson and colleagues (1992). This amounted to the satisfaction of the following constraints: First, reinterpretation rates were required to be lower for abstract, conceptual interpretations than for geometric interpretations (cf. Finke et al., 1989). In addition, interpretations obtained during mental imagery had to be below those obtained during visual perception.

Second, reinterpretation rates were required to be qualitatively consistent with the findings of Peterson and colleagues (1992), which indicate that reinterpretation rates increase after a de- and refocus of attention.

Calculation of parameter effects

The calculation of individual parameter effects and parameter interactions was based on a design matrix of –s and +s, representing high- and low parameter values (cf. figure 1). In this matrix each column denoted a model parameter and each row represented a specific parameter combination. A measure of model performance, that is simulated mental reinterpretation probability, was associated with each row in the design matrix. In general, in order to obtain a parameter's average effect on overall model performance, those rows in the model performance column of the design matrix which correspond to a low parameter value are summed and subtracted from those rows which correspond to high values. Higher-order interaction effects can be obtained in a similar manner (Law and Kelton, 1991; Box and Hunter, 1978). Given the simulation design matrix, these calculations can be expressed as a sequence of simple matrix operations.

High-dimensional visualizations

Those groups of interacting parameters whose modulating effect exceeded 20% of the central parameter's effect—in our project this parameter denoted the focusing of attention—were prepared for subsequent visualization. Simulation data was prepared in such a way that parameters which exhibited a stronger mutual interaction with the central parameter would also be visualized closer to each other. This grouping of more related parameters turned out to enhance the understanding of interactions, since stronger interaction patterns emerged as salient color-patches.

The visualizations (illustrated in figure 3) can be conceived of as a high-dimensional cube of changes in model performance, each dimension representing changes caused by one of the interacting parameters. This cube can be sliced and stacked recursively onto a two-dimensional plot (cf. Bosan and Harris, 1996; Harris et al., 1994). Each x-y coordinate in these plots denotes a specific combination of interacting parameters. In our project, the direction of change in model performance was coded along two different color scales, and the magnitude of change was indicated by variations in hue within these scales, with deeper colors depicting a larger change.

The amount of information contained in the visualizations was further increased by the addition of information on model validity. We let the relative area of each colored square reflect the average validity of models corresponding to the central parameter's high value. In our case, this amounted to selective attention being focused. As a result of including model validity in the visualizations, simulation data contributed to the visual appearance of the plot only to the extent to which they were valid.

What type of results can be obtained?

Two categories of questions can be addressed using this method. First, simulation results can be approached with a particular hypothesis in mind, as was done in our project. In this case, one would like to make sure that the main effect of a particular embodiment of a cognitive mechanism, x_+ (corresponding to parameter x at its high value), is as was predicted. For example: Do any of the interactions observed in the simulation results change the fact that parameter x is generally inhibitory? In addition, one would be interested in mapping out the validity of models where cognitive mechanism x_+ is operating.

Second, simulation results can be openly explored, perhaps focusing on the role of a few central parameters. In this situation, one could, for example, be interested in finding out which cognitive mechanisms work in concert and which work against each other. In the first case the mechanisms would affect model performance in the same direction. In the latter case they would work in opposite direction, canceling out each other's effect. In addition to mapping out such interactions, one would be interested in which combination of mechanisms constitute valid models. This search for valid models can be a powerful way of constraining the space of possible models when several sources for validation are used (for example, a small set of seemingly contradictory experimental results).

Concluding discussion

The use of distinctive colors, the organization of the visualizations' layout according to the strength of interactions, together with the technique described above for indicating model validity, turned out in practice to facilitate the understanding of the interaction patterns. Strong interactions which also gave rise to valid performance tended to visually coagulate into contiguous color-patches, which "popped-out" from the background of empty squares, marking nonvalid cases.

The virtues of this combination of factorial simulation, analysis and visualization method are, in our view, compelling: Although the modeling framework is assumed to be based on a firm empirical basis, model properties which are not well-founded need not be specified in an ad hoc manner.

From a more theoretical perspective, conclusions which can be drawn from a full-factorial investigation will approach the stringency of appropriately conducted "real-world" experiments, with an inevitable difference: The validity of any results obtained will ultimately depend on the validity of the modeling framework itself. Within this framework, causal dependencies between hypothetical cognitive mechanisms and overall model behavior can be correctly mapped out. As a result, the development of subsequent models and/or the construction of cognitive theories can be guided in a stringent way.

As the method itself is qualitative in nature (parameters are varied coarsely between a high and a low value), models can be validated on the basis of qualitative empirical data. Note that the objective with using this method is not primarily to quantitatively adjust a model's overt performance to empirical data by manually tuning parameters, but instead to single out a combination of internal cognitive mechanisms as the probable cause of empirically observed human behavior.

In a longer perspective, this method can contribute to the incremental development of more and more finely tuned cognitive models. Starting with a firmly based, minimally specified initial model framework, valid cognitive mechanisms can be singled out and subsequently embedded into the framework. Given these additional mechanisms, and/or having refuted some peripheral model properties, the next round of search can be narrowed down, and targeted at a more detailed level. As each increment is reasonably well-founded (validation is based on average simulation results), model development can be more directed.

Acknowledgments

We would like to thank Sture Hägglund, and two anonymous reviewers for suggestions and valuable comments on an earlier version of this article. We would also like to thank Yvonne Wærn, and Jonas Barklund for fruitful discussions during the development of this project. This work was supported by the Swedish National Board for Industrial and Technical Development and the Swedish Council for Research in the Humanities and Social Sciences.

References

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Lawrence Erlbaum.

Bosan, S. & Harris, T. R. (1996). A visualization-based analysis method for multiparameter models of capillary tissue-exchange. *Annals of Biomedical Engineering, 24,* 124-138.

Box, G. E. P., Hunter, W. G., & J. S. (1978). *Statistics for experimenters: An introduction design, data analysis, and model building.* New York: Wiley.

Finke, R. A., Pinker, S. & Farah, M. J. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science, 13,* 51-78.

Harris, P. A., Sorel, B., Harris, T. R., Laughlin, H. & Overholser, K. A. (1994). Parameter identification in coronary pressure flow models: A graphical approach. *Annals of Biomedical Engineering, 22,* 622-637.

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

Kieras, D. E. (1987). Cognitive modeling. In Shapiro, S. C & Eckroth, D. (eds): *Encyclopedia of artificial intelligence, vol 1.* New York: Wiley.

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1994). *Image and Brain: The resolution of the imagery debate.* Cambridge, MA: MIT Press.

Kosslyn, S. M., Pinker, S., Smith, G. E. & Swartz, S. P. (1979). On the demystification of mental imagery. *The Behavioral and Brain Sciences, 2,* 535-581.

Kosslyn, S. M., Flynn, R. A., Amsterdam, J. B., Wang, G. (1990). Components of high-level vision: A cognitive neuroscience analysis and accounts of neurological syndromes. *Cognition, 34,* 203-277.

Kovordányi, R. (1998). Is mental imagery symbolic? Exploratory simulations in an interactive activation model. In *Proceedings of Second European Conference on Cognitive Modeling.* Nottingham: Nottingham University Press.

Kovordányi, R. (1999a). Mental image reinterpretation in the intersection of conceptual and visual constraints. In Paton, R. & Neilson, I. (eds): *Visual representations and interpretation.* London: Springer Verlag.

Kovordányi, R. (1999b). Modeling and simulating inhibitory mechanisms in mental image reinterpretation—Towards cooperative human-computer creativity. *Linköping Studies in Science and Technology.* Dissertation no. 589. ISBN 91-7219-506-1. Linköping: Linköping University Press.

Kovordányi, R. (2000). Full factorial simulation modeling of selective attention in mental imagery. Presented at *XXVII International Congress on Psychology,* Stockholm.

Law, A. M. & Kelton, W. D. (1991). *Simulation modeling and analysis.* New York: McGraw-Hill.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 4, 287-330.

McClelland, J. L. & Rumelhart. D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 5, 375-407.

McClelland, J. L. & Rumelhart. D. E. (1994/1988*). Explorations in parallel distributed processing: A handbook of models, programs and exercises.* Cambridge, MA: MIT Press.

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

Rosenbloom, P. S., Laird, J. E., & Newell, A. (1993). *The Soar papers.* Cambridge, MA: MIT Press.

Rumelhart. D. E. & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 1, 60-94.

Peterson, M. A., Kihlstrom, J. F., Rose, P. M. & Glisky M. L. (1992). Mental images can be ambiguous: Reconstruals and reference-frame reversals. *Memory and Cognition, 20,* 107-123.