Factorial Modeling: A Method for Enhancing the Explanatory and Predictive Power of Cognitive Models

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

Abstract

The construction and evaluation of cognitive models can, and often do, lead to *novel* insights into what might constitute a valid account for an empirical phenomenon. These insights constrain the space of viable models, and could be useful also on a theoretical plane, by promoting a deeper understanding of the studied phenomenon. We propose the factorial method for deriving novel, that is, not theory–based constraints *in a principled way* during model development. The method is based on a systematic comparison of alternative models, realized through a cross–combination of model components in a generic cognitive model. We illustrate the method by describing an application in the area of mental imagery. We conclude by discussing ways to increase the generalizability of results that can be obtained using the factorial method.

Introduction

From a modeling perspective, cognitive theories are presumed to guide and constrain model construction. Often however, there is a considerable gap between what has been theoretically established and what can be consistently modeled and simulated. This gap may not only concern an inherent difference in levels of description, but could additionally reflect a genuine lack of knowledge about the studied phenomenon. Bridging this gap is thus a nontrivial task: It often requires numerous iterations between tentative model construction and evaluation.

In this process, new insights may emerge regarding how cognitive models ought to be constructed in order to fit the empirical data. From the modeler's perspective, these insights constrain the space of viable models, further narrowing down the envelope allowed by the underlying cognitive theory. The theoretical value of these constraints will in part be determined by their original motivation: whether they are motivated by implementational considerations or have a logical basis (cf., e.g., Cooper et al, 1996).

Additionally, the theoretical value of newly discovered constraints may also depend on whether they were accidentally found, and thus cannot be guaranteed to hold in all cases, or if they were systematically uncovered. Note that the intrinsic requirement for internal consistency and computational tractability that computational models must comply with, and the stringency that these requirements impose on model development, would provide a firm basis for deriving non-theory-based constraints—as long as alternative models are evaluated and compared in a principled way.

As an example of *systematic* exploration of alternative model solutions, Kieras and Meyer (1995) describe an investigation where dual–task performance was modeled using EPIC, a symbolic unified cognitive architecture. Various resource–sharing strategies were explored and the corresponding reaction times simulated. Simulated reaction time for alternative strategies were compared to empirical data on dual–task performance. On the basis of an extensive search for alternative strategies which would reproduce the empirical data, the authors draw the conclusion that human subjects must be using a near–optimal task strategy, pipelining their visual input for one task, while executing the other task. This conclusion is based on the fact that the authors were not able to fit model performance to the empirical data using any other strategy, given the framework of EPIC.

The question of whether and how a search for feasible model properties should be conducted is common for many modeling projects. Model development and validation often involves a more or less systematic search for model properties (parameter values) that make the model behave in the desired way and reproduce the empirical data. What is important to realize in this context however, is that individual model properties may be dependent on each other. In other words, one model property may affect model validity in a certain way only when other model components are present, or are implemented in a certain way. In this situation, the model may only reproduce empirical data if a particular *combination of model properties* is present.

We propose a formal (and automatic) method for mapping out the intrinsic dependencies between model properties, while also estimating their individual contribution to model performance. This method relies on a systematic exchange of model components and/or alternative implementations of model components, and an evaluation of their effect on model validity or some other measure of model performance.

In a broader perspective, the proposed method entails a shift of focus from simply demonstrating *that* a specific model is valid, to characterizing *under which conditions* the model—or rather a generic model framework—could account for empirical data. In this sense, our proposal could be seen as a first step towards a more principled way of theory testing (cf. Roberts and Pashler, 2000).

In the following sections, we will shortly describe the factorial method, and illustrate its use by accounting for an example application in the area of mental imagery (Kovordányi, 1999, 2000a). Finally, we will discuss limitations in the generalizability of results obtained with the method, and propose an extension of the method as a way of dealing with these limitations.

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 some "reasonable value" in order to keep down simulation complexity, the power of the simulation design would decrease. Simply expressed, parameters may have been fixed at a value where they strongly modulate the effect of central model parameters.

In practice, a minimal set of model properties will inevitably be determined a priori on the basis of the underlying cognitive theory. This generic model framework could still leave unconstrained a large number of model design decisions. How should the final simulations be designed if the corresponding number of model parameters turn out to be unmanageably large?

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 used instead of a full design (cf. Law and Kelton, 1991; Box et al, 1978). Note that in these designs, peripheral parameters are not fixed at an ad hoc value, but are instead defined dynamically as a function of those parameters which are varied.

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 conceptually 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 (cf. figure 1). 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 data analysis procedure, let us assume that the possible interaction between parameters p_1, p_3, and p_7 are inquired. In this case, columns 1, 3, and 7 of the design matrix are multiplied with each other entry-by-entry, and then multiplied, again entry-by-entry, with the corresponding simulation results. The effect of these multiplications is that the correct signs will be added to the results–column. A final summation of all the signed entries in the results–column, divided by 2^{k-1}, where k denotes the number of model parameters varied, yields the desired mean interaction of the parameters involved.

run	par 1	par 2	sim. result
1	–	–	R_1
2	–	+	R_2
3	+	–	R_3
4	+	+	R_4

Figure 1: A two-level factorial design matrix for two 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.

Application of the method

In the following sections, we will briefly describe an investigation of mental imagery where a full two-level factorial design was used (Kovordányi, 1999, 2000b). Although the effect of several possible factors, such as mental image fading, were taken into account, the analysis of simulation results was centered on revealing 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), which were used for model validation, were qualitative, no attempt was made to optimize the models towards these data. Model validity was instead defined qualitatively, and served as a means for evaluating the feasibility of alternative models.

Identifying variable model components

The model framework used in our project drew its main architectural components from the comprehensive model of mental imagery developed by Kosslyn (1994; Kosslyn et al, 1979; Kosslyn et al, 1990). Within this framework, lower–level model components remained partially unconstrained. For instance, should attentional selection be implemented as an early or late selectional mechanism? Is selective attention involved (focused) at

all during mental image reinterpretation? These choices were expressed as variable model components that were systematically exchanged between simulation runs to allow for a comparison of various model instances. As a result, half of all simulation runs would be based on models containing a late selectional model component, a quarter of all simulations run would be based on models containing a late selectional component *and* also implementing an inhibitory fringe around the selectional 'spotlight', etcetera.

We chose to implement our model framework as an interactive activation model (cf. McClelland, 1979; McClelland and Rumelhart, 1981, 1994/1988;). In these models, the localist nodes are arranged into reciprocally connected layers of processing, thereby increasing the structure and penetrability of the model. Units within the same processing layer are assumed to have the same inhibitory and excitatory connection weights.

Within our interactive activation model framework, variable model components were expressed in terms of connection weights, activation thresholds, resting levels, and/or "control flags". Control flags were also used to control whether processing was to be initiated top-down or bottom-up. These two modes of processing corresponded to mental imagery vs. visual perception in human subjects.

Variable model components could equally well be delineated in symbolic models, as alternative (sets of) production rules, or simply alternative definitions (fnc1 – fnc2) of a cognitive mechanism together with a means for activating them at run-time. Hence, the factorial method can be applied to any modularly constructed computational model with a minimal overhead cost.

Simulations

Our model framework for mental imagery encompassed three mutually interacting layers of processing (figure 2). At the lowest level, the visual buffer contained detectors for oriented line segments. At the next stage, these feature detectors would evoke (and receive feedback from) simple geometric patterns, such as composite lines and triangles. These patterns were stored in visual long-term memory. At the highest level of processing, the low–level geometric patterns were combined into concepts stored in associative long-term memory. In addition to the between-layer connections, we assumed lateral that is, within-processing-level inhibition, between mutually inconsistent groups of computational units. Image interpretation in this cascading system was based on the establishment of a correspondence between lower-level and higher-level representations across the processing layers.

We simulated mentally- and perceptually based reinterpretation of two composite line drawings adopted 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' superimposed 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".

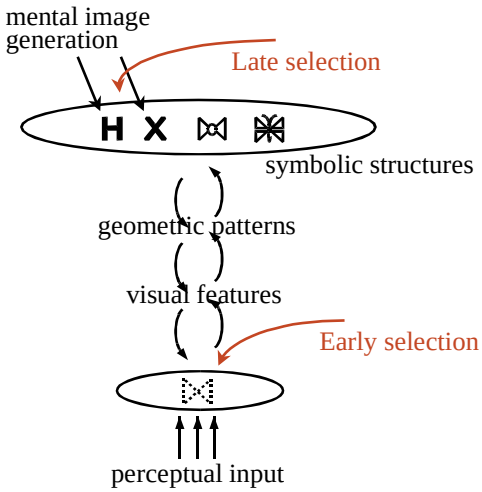

Figure 2: An outline of the interactive activation model used in our simulations of mental and perceptual image reinterpretation. Processing is based on three bi–directionally connected layers of localist units. Processing in cascade allows focus of attention to be propagated with a negligible time delay to both earlier and later stages of processing. In spite of this, the point of initiation of attentional focus turns out to influence model validity.

As processing layers were reciprocally connected, 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 perceptual stabilization of the same line-figure, concluded by perceptually based reinterpretation. Each simulation was run for 10 simulated seconds in discrete steps of 50 ms.

Two instances of the model framework were scrutinized: One where attentional selection affected processing at a late stage, at the level of associative long-term memory, and one where selection was initiated early, at the level of the visual buffer. For these models,

the effect of focusing attention (versus not focusing attention) was investigated, taking into account that interaction might arise between these central and other peripheral model components.

Data analysis

In this example project, data analysis began with semi-automatic preparation of the raw simulation data (see below). 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 model components

For simplicity, we will denote model components as *simulation parameters* in the sections on data analysis. Activation levels of all response units in the interactive activation network were measured for each simulation run that is, for each parameter combination. From these activation values the probability for reinterpretation was calculated. Reinterpretation rates were classified as valid if they qualitatively matched the reinterpretation rates obtained by Finke and colleagues (1989, exp. 1), and Peterson and colleagues (1992).

These empirical data posed the following constraints on the simulation results: First, reinterpretation rates were required to be less for symbolic than for geometric interpretations (cf. Finke et al, 1989). In addition, interpretations obtained during mental imagery had to be below those obtained during the perceptual phases of the simulations.

Second, reinterpretation rates were required to be qualitatively consistent with the findings of Peterson and colleagues (1992). These findings are interpreted as an indication that reinterpretation rates should increase after a de- and refocus of attention.

Calculation of model component effects

The calculation of individual component effects and interactions was based on a design matrix of –s and +s, representing high- and low simulation parameter values (cf. figure 1). In this matrix each column denoted a model parameter and each row represented a specific parameter combination. Two measures of model performance: simulated mental reinterpretation probability and model validity, were 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 one of the results–column which corresponded to a low parameter value were summed and subtracted from those rows which corresponded to high values. Higher-order interaction effects were obtained in a similar manner (Law and Kelton, 1991; Box and Hunter, 1978). Given the simulation design matrix, these calculations could be expressed as a sequence of simple matrix operations.

Visualizations of interactions

Those groups of interacting parameters whose modulating effect exceeded 20% of the central parameter's effect (in our case this parameter denoted the focusing of attention) were prepared for visualization.

The type of visualizations obtained (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).

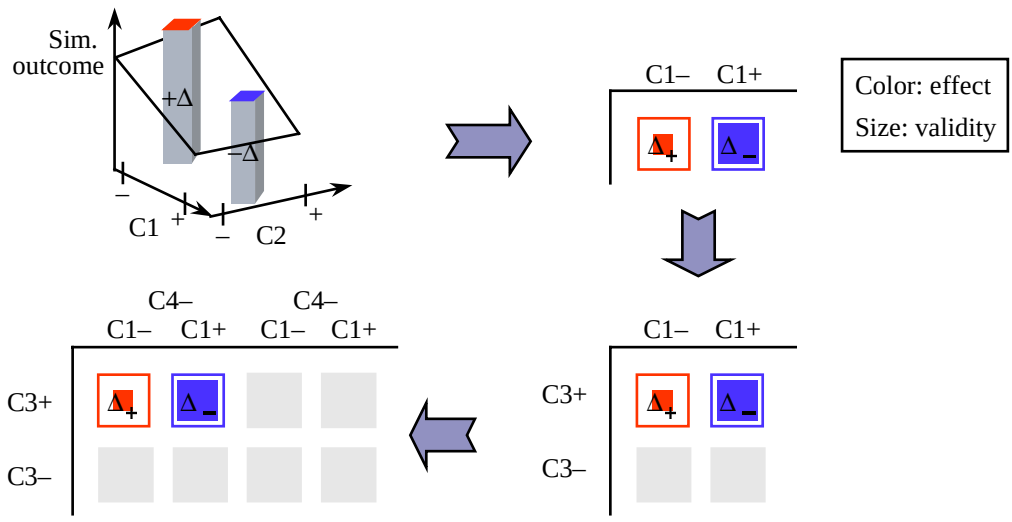

Figure 3: Visualization of the simulation results is achieved through recursive slicing of the high–dimensional volume of simulation data. Simulation results are color-coded to facilitate the perception of interaction patterns. The relative area of the square markers reflects mean model validity for the underlying parameter combinations.

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 darker, more saturated colors depicting a bigger change.

In addition, we made 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 during image reinterpretation. 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.

Results obtained in the example project

We focused our simulation and data analysis on the question of early vs. late attentional selection, our hypothesis being that late selection would account for the empirical data, while early selection would not. This hypothesis seems to be supported by our simulation results: In short, models were valid when selective attention was focused during image reinterpretation—as compared to simulations when attention was not focused. In addition, when comparing models containing an early vs. late selectional component, the latter models turned out to produce valid behavior, while the former did not. Although these results can be interpreted an indication of an overall pattern, the generalizability of these results need to be further examined.

Extension of the method

By its systematicity, the factorial method enhances the reliability of any constraints discovered during model development and simulation. However, we would like to point to one limitation of this method. The factorial method, in the form presented above, is aimed at characterizing the space of alternative models. The underlying assumption is that transitions in this space are smooth that is, slow, and monotonic, and hence can be characterized on the basis of the two data points per dimension used to calculate parameter effects.

We see two problems with this assumption. First, model validity, or model performance in general, could vary between the sampled points. Second, model performance is characterized on the basis of a limited subspace of the complete model space. For example, connection weights in a connectionist model might have been varied within a narrow range, which may not cover the complete interval allowed for that particular type of connection. Hence, component effects and interactions might look different both within and outside the subspace, or *segment*, which has been sampled. Both of these limitations affect the generalizability of results obtained using the factorial method

Hypothesis testing by searching for counterexamples in model space

We would like to propose one way of approaching this problem. For practical reasons, we cannot ensure in the general case that any effects found will hold throughout model space. However, researchers are frequently interested in finding support for or refuting one particular hypothesis. For example, it would be interesting to know if an assumption of late attentional selection is *the only way* to account for empirical data. This amounts to the question of 'Would early attentional selection account for empirical data if a different segment of model space was examined?'

In this limited setting, a search of model space becomes tractable. The objective is to examine various segments in model space, in order to ensure that any results found in one segment are general enough to also hold in other segments of the model space. This extension of the method relies on *random* sampling of segments using, for example, genetic algorithms.

Note that while genetic search can be used to delimit various segments in model space, characterization of each of these segments must be based on the factorial method. The reason for this is that in order to be able to attribute model validity to a specific model component that is, exclude the possibility of some peripheral aspect of the model affecting model validity, all variable components must be cross–combined. In essence, we want to detect *causal relationships* between model component(s) and model validity.

Hence, for example, the objective in the example project would be to ensure that late selection *causes* models to be valid in all segments in model space. In other words, we want to ensure that model validity can be attributed to late selection, and not some fortunate interaction of other model components.

In the empirical sciences, a favorite hypothesis is supported by evidence, when the scientist has done everything to prove its negation, the null hypothesis, and failed. In the same manner, the objective in the example project could be to search for segments in model space where model validity can be attributed to, not late, but *early* selection.

There can be two outcomes of such a search. The first possibility is that early selection models turn out to be invalid throughout model space. This result could be used as a basis for making general statements about the necessity of a late selectional mechanism in models of mental imagery. The second possibility is that early selection turns out to result in valid models in some segments of model space. In this latter case, one might attempt to detect common features in those segments where early selection turned out to result in valid models. Again, this would produce generalizable *new* knowledge about the studied phenomenon.

Summary

As is often pointed out in the modeling methodology literature (cf., e.g., Cooper et al, 1996), there is an inherent gap between cognitive theories and their realizations as computable cognitive models. Novel, that is, not theory–based constraints on what could constitute a viable account for an empirical phenomenon are thus often discovered during the development and testing of cognitive models. These constraints could turn out to be *theoretically* useful, provided that they were uncovered in a systematic fashion.

We propose the factorial method for deriving novel that is, not theory–based, constraints in a principled way. The method relies on a systematic validation and comparison of alternative models, and in practice, entails a shift of focus from simply demonstrating *that* a specific model is valid, to characterizing *under which conditions* the model can account for empirical data.

The method provides a formal basis for stating that, given a set of fundamental, theory–based assumptions, the studied phenomenon can be modeled successfully only if certain additional assumptions are made. These assumptions can be about subjects' choice of task strategy when performing dual–tasks, or concern the necessity of a particular cognitive mechanism in models of mental imagery.

The reliability of model constraints is increased if the causal relationship between the inclusion of a specific model component and resulting model validity can be demonstrated to hold irrespective of which part of model space is examined. As model space cannot, in general, be searched in its entirety, we suggest a more focused approach of hypothesis testing: Given an initial hypothesis, model space is searched for sub–segments in which a designated alternative model solution leads to valid models. Depending on the outcome, the initial hypothesis can be reliably refuted or supported.

Acknowledgments

We would like to thank three anonymous reviewers for valuable comments. This work was supported by the Swedish Council for Research in the Humanities and Social Sciences.

References

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.

Cooper, R., Fox, J., Farringdon, J., & Shallice, T. (1996). A systematic approach for cognitive modeling. Artificial Intelligence, 85, 3-44.

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., & Meyer, D. E. (1995). Predicting performance in dual-task tracking and decision making with EPIC computational models. In *Proceedings of the First International Symposium on Command and Control Research and Technology*.

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. (1999). 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. (2000a). Full factorial simulation modeling of selective attention in mental imagery. Presented at *the Twenty Seventh International Congress on Psychology,* Stockholm.

Kovordányi, R. (2000b). Controlled exploration of alternative mechanisms in cognitive modeling. In *Proceedings of the Twenty Second Annual Meeting of the Cognitive Science Society*.

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.

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.

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