Modeling sequential information integration with parallel constraint satisfaction

Katja Mehlhorn (katja.mehlhorn@phil.tu-chemnitz.de)
Chemnitz University of Technology, Department of Psychology
Georg Jahn (georg.jahn@uni-greifswald.de)
University of Greifswald, Department of Psychology

Abstract

An important aspect of human cognition is the sequential integration of observations while striving for a consistent mental representation. Recent research consistently stresses the importance of fast automatic processes for integrating information available at a certain point in time. However, it is not clear, how such processes allow for maintaining a consistent and up to date mental representation in the light of new information. We compare variants of two methods of modeling sequential information integration with parallel constraint satisfaction models: carrying over results from the previous integration step or decaying input strength of older observations. Results of these models for consistent and inconsistent sets of observations are compared to human data from a diagnostic reasoning task.

Keywords: information integration, belief updating, diagnostic reasoning, memory activation, constraint satisfaction modeling

Introduction

A key feature of many everyday reasoning tasks is that observations are processed sequentially. Whether it is in diagnostic reasoning, in decision making, or in belief updating, often information becomes available step by step. If a large amount of information is given all at once, it might only be perceived and understood sequentially due to limited cognitive capacities. Although possible implications of the sequential nature of tasks (e.g., order effects) have been discussed (e.g., Hogarth & Einhorn, 1992; Wang, Johnson, & Zhang, 2006), the underlying cognitive mechanisms are not fully understood. Recent research consistently points out the importance of fast automatic processes for integrating information available at a certain point in time (e.g. Glöckner & Betsch, 2008). However, it is not clear how such processes allow for maintaining a consistent mental representation in the light of new incoming information. In this paper, we explore alternative implementations of such processes in connectionist constraint satisfaction models.

Previous research has shown that reasoners hold knowledge structures that reflect the structure of the task in the environment (e.g., Anderson, 1983; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Thomas, Dougherty, Sprenger, & Harbison, 2008). For example, a physician learns, with an increasing number of patients encountered, which symptoms are associated with which diseases and how strong these associations are. Given such an adapted knowledge structure, observations can serve as a cue for the retrieval of associated knowledge from long-term memory (e.g., Kintsch, 1998; Thomas et al., 2008; Baumann, Mehlhorn, & Bocklisch, 2007). To maintain a consistent representation of the task at hand, this newly activated information somehow needs to be integrated with previous observations and previously activated knowledge. How is this achieved?

Wang et al. (2006) have proposed a connectionist model of sequential integration based on the idea of explanatory coherence that, probably most prominently, was introduced by Thagard (1989, 2000) in the field of scientific discovery. Thagard implemented explanatory coherence among interconnected propositions in a connectionist constraint satisfaction model (ECHO). In ECHO, propositions are represented by nodes. The nodes are interconnected by symmetric excitatory and inhibitory links representing the relations (constraints) between them. Nodes representing observed information are additionally connected to a special activation node (special evidence unit = SEU), which always has an activation value of 1 and is the model's "energy source". Connecting not all, but only these data nodes to the energy source reflects the idea that empirical data are weighted more strongly than theoretical hypotheses held by the reasoner (Thagard, 1989).

The strength of a proposition in the network is indicated by the numerical activation of its node. Before the network is integrated, activation of all nodes is set to default values. Then, activation spreads from the SEU to the data nodes and then to other connected nodes. The net input each node receives is calculated as the weighted sum of the activation of all nodes it is connected to. After calculating the input for each node, the activation of all nodes is updated synchronously. These two steps are repeated iteratively, until activation stops changing substantially. The more consistent a proposition is with the observed information and other related propositions, the higher is the activation of its node when the network settles.

The idea of constraint satisfaction has been widely applied to areas such as text comprehension (Kintsch, 1998), social impression formation (Thagard & Kunda, 1998), visuo-spatial reasoning (Thagard & Shelley, 1997), causal reasoning (Hagmeyer & Waldmann, 2002), medical diagnosis (Arocha & Patel, 1995), and decision making (Glöckner & Betsch, 2008). In all of these different tasks, reasoners need to find an interpretation that is more coherent with the available information than possible alternative interpretations. Such coherent interpretations can be the meaning of a word that fits best in the current context, the impression about a person that is most coherent with one's previous impression about him/her, or it can be the diagnosis that best explains the set of a patient's symptoms.

Applied successfully to model various phenomena in all the above domains, constraint satisfaction models have been described as a "computationally efficient approximation to probabilistic reasoning" (Thagard, 2000, p. 95). However, Thagard's ECHO has some major limitations. For our question most importantly, it only models the parallel integration of information given at a certain point of time. To incorporate newly incoming observations in a sequential task, a new network would have to be constructed.

Wang et al.'s UECHO (uncertainty-aware ECHO; 2006), shares the basic features of ECHO, but can handle sequentially incoming observations. This is achieved by two basic changes. First, the network contains not only the currently available information as in ECHO, but all possible observations are included from the beginning. Thus, when new observations come in, the network does not have to be restructured. Second, the models differ with regard to which observations are connected to the special evidence unit (SEU). While in ECHO, all observation-nodes are connected to the SEU, in UECHO, only those nodes representing information observed until the current point of time are connected to the SEU. Due to these two changes, when a new piece of information is observed, the model does not have to be rebuilt, but only a new connection between that observation and the SEU needs to be added.

For modeling sequential information integration, it is not only important to incorporate new observations into the network, but also to consistently integrate this new information with the previous state of the network. One could think of two basic approaches for implementing this preservation of the previous state (shown in the models in Figure 1). In both models, the upper nodes, E1 and E2, represent possible explanations of the observed symptoms S1-S4 (represented by the nodes in the middle row). Solid lines between the nodes represent coherent relations (e.g., E1 explains S1), dashed lines represent incoherent relations (e.g., E1 and E2 contradict each other). In both models, the symptoms S3 and S1 have been observed.

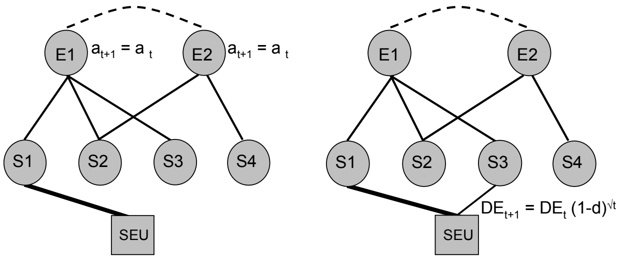

Figure 1: Two basic approaches to model sequential data in a constraint-satisfaction network. Either the previous state of the model is preserved by retaining the initial activation of the explanation nodes (left) or previous symptoms keep influencing the activation in the network by a (decaying) connection to the SEU (right).

In the left model, the previous state of the network is preserved by retaining the activation of the explanation nodes. After the first symptom (S3) is observed, the activation for the explanation nodes is calculated. This activation is then used as starting value for the integration of the new symptom (S1).

On the right, the approach proposed by Wang et al. (2006) is illustrated. Here, activation is reset to default before each new run. The preservation of the previous state is obtained indirectly, by connecting not only the new information, but also previously observed information to the SEU. In the model, S3 as well as S1 is connected to the SEU. Therewith, the older observation can continue influencing the current activation in the network. To account for sequential observations, the strength of this influence decays over time. The most recently observed symptom (S1) gets a strong connection to the SEU, whereas older observations are connected to the SEU with a decayed strength (S3). This strength (data excitation - DE) is a function of a decay rate *d* and the time interval since the symptom was observed. By referring to work on memory retention, Wang et al. (2006) propose to let DE decay exponentially in the square root of time.

We will show that the first modeling alternative - retaining output activation from previous runs - is not appropriate for modeling the integration of sequential information, because of the dynamics of spreading activation. The second alternative is explored in more detail. The resulting activation for both approaches is tested against human data.

Experiments

Design and procedure

Human data on memory activation during sequential symptom integration was obtained in two diagnostic reasoning experiments (see also Baumann et al., 2007; Mehlhorn, Baumann, & Bocklisch, 2008). In these experiments, participants were to diagnose hypothetical patients after a chemical accident. For each patient, a set of symptoms was presented sequentially on a computer screen and the task was to find the chemical that best explained this set of symptoms. The knowledge necessary to solve this task was taught to participants in an extensive training session. The knowledge consisted of nine different chemicals (named with single letters), grouped into three categories (see Table 1).

Table 1: Domain knowledge participants had to acquire before Experiment 2 (original material in German).

Category	Chemical	Symptoms
Landin	B	cough, short breath, headache
	T	cough, vomiting, headache, itching
	W	cough, eye inflammation, itching
Amid	Q	skin irritation, redness, headache
	M	skin irritation, short breath, headache, itching
	G	skin irritation, eye inflammation, itching
Fenton	K	diarrhea, vomiting, headache
	H	diarrhea, redness, headache, itching
	P	diarrhea, eye inflammation, itching

Each chemical caused three to four symptoms. Symptoms were ambiguous, as each symptom could be caused by two to six different chemicals. So, only the combination of symptoms allowed for unambiguously identifying the correct diagnosis.

Two types of trials were used in the experiments; consistent and inconsistent trials (see Figure 2). In consistent trials, all symptoms consistently pointed towards one explanation. Thus, the participants' initial explanation was supported by all later symptoms. In inconsistent trials, the explanation suggested by the first two symptoms was inconsistent with the later symptoms. Here, participants needed to revise their initial explanation after observing the third symptom. In such inconsistent trials, it should be particularly difficult to integrate symptoms while maintaining a consistent mental representation. In Experiment 1 (Baumann et al., 2007), participants were presented with a total of 340 consistent trials. In Experiment 2 (Mehlhorn et al., 2008), participants worked through a total of 384 trials, of which 75% were consistent and 25% were inconsistent.

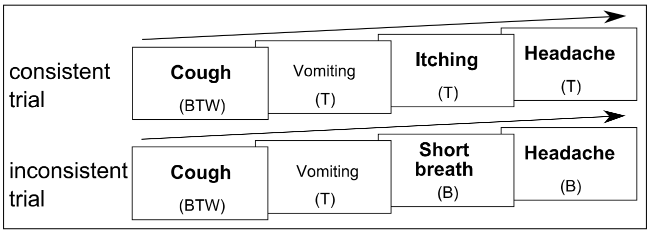

Figure 2. Example for a consistent and an inconsistent trial in Experiment 2. Letters in parentheses represent explanations consistent to all previous symptoms.

In both experiments, two types of dependent measures were obtained. First, after all symptoms of a patient were presented, participants explicitly provided their diagnosis. Second, a probe reaction task was used as an implicit measure of the activation of explanations during the sequential task. This measure is based on the idea of lexical decision tasks (e.g., Meyer & Schvaneveldt, 1971) according to which participants should respond faster to a probe that is higher activated in memory than to a probe of low activation. Each probe was a single letter that was either one of the names of the nine chemicals (explanations) or one of nine other letters. Participants were to decide as fast as possible whether the probe was a chemical name or not. To reduce possible influences of the probes on each other, only one single probe was presented in each trial. Using this measure, it was possible to monitor the activation of explanations in the course of the sequential reasoning task with as little impact on the task itself as possible.

Such an implicit measure that directly tracks the activation of explanations in memory is especially suited to evaluate the validity of constraint satisfaction models. The usual approach to test these models is to compare the activation calculated in the model to an explicit measure obtained in human experiments. For example, Wang et al. (2006) asked their participants for explicit belief ratings after each new observation. However, explicit belief ratings have a major drawback. Asking participants during the course of the task might influence the outcome of the task itself (c.f. Hogarth & Einhorn, 1992). Directly assessing the activation in memory with an implicit task is less reactive.

In this paper, we use response time data for three different types of explanations to assess constraint satisfaction models. First, we are interested in explanations that are most consistent with all symptoms observed before the probe's presentation (*relevant* explanations). Second, we are interested in explanations that participants considered relevant after earlier symptoms, but are inconsistent with later symptoms (*rejected* explanations). Third, we look at explanations that were inconsistent already with the first symptom of the trial (*irrelevant* explanations).

Results

Diagnosis In both experiments and in both types of trials, the accuracy of diagnoses given at the end of each trial was quite high (around 95%). This suggests that also in inconsistent trials, participants were able to solve the task easily.

Probe Reaction Task The fastest responses in the probe task occurred for explanations that were relevant given the symptoms observed up to then. Rejected explanations were responded to slower than relevant explanations, but faster than irrelevant explanations. This basic pattern was found in consistent trials in both experiments, as well as in the inconsistent trials in the second experiment (see Figure 3). However, consistent and inconsistent trials differed in the courses of activation over time. In consistent trials, reaction times decreased with increasing number of symptoms. In inconsistent trials, this decrease was less visible, as integrating the information was more difficult than in consistent trials. Nevertheless, the fast responses to relevant explanations show that participants managed to integrate the symptoms properly.

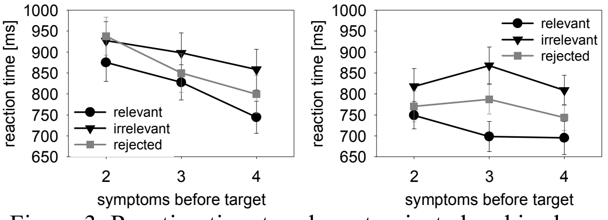

Figure 3. Reaction time to relevant, rejected and irrelevant probes after the 2nd, 3rd, and 4th symptom. Left graph: consistent trials (Baumann et al., 2007); Right graph: inconsistent trials (Mehlhorn et al., 2008)

Models

To assess the validity of the alternative modeling approaches, we implemented the knowledge used in the experiments into different constraint-satisfaction networks. All networks consisted of the whole material participants needed to learn before the experiment (see Figure 4). We used 9 nodes representing the symptoms, 9 nodes representing the explanations (chemicals), and several connections representing the relations between those nodes. Nodes representing explanations were interconnected by inhibitory links, because the symptoms of each trial were caused by only one chemical. Explanations and symptoms were interconnected by excitatory links.

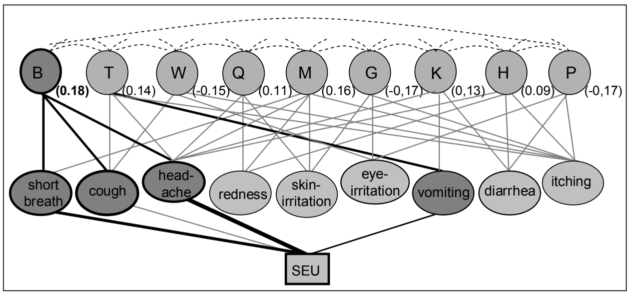

Figure 4. Network for an inconsistent trial in Simulation 3. Inhibitory connections: dashed lines, excitatory connections: solid lines. Numbers in parentheses: activation values of explanation-nodes after the network settles.

In the networks, four basic parameters can be varied.

1. The initial activation of the explanation-nodes before each run.
2. The initial activation of the symptom-nodes before each run.
3. The strength of the connection between the nodes.
4. The strength of the connection between the symptom nodes and the special evidence unit (SEU).

To model the two basic approaches described above, we used variations of the parameters 1 and 4. The results for these 2 approaches are presented below. For both alternatives, we ran various models, testing different values for parameters 2 and 3. These values did not have any substantial effect on the model outcome. Therefore, in the models described below, they are set to fixed values. The initial activation of symptom nodes (parameter 2) is set to 1 for the currently observed symptom and to 0 for all other symptoms. The connection-strength between nodes in the network (parameter 3) is set to 0.04 for excitatory and to -0.04 for inhibitory links.

To evaluate the models' capacity to emulate human information integration during the course of the task, we will now take a closer look at the process measure. For each model, we calculated the activation for the three types of explanations (relevant, rejected and irrelevant) at three different times of measurement (after 2, 3, or 4 symptoms). This activation is directly compared to the human response time data, which indicate memory activation of explanations.

Initial activation of the explanation-nodes

Simulation 1 and 2 One method to model sequential data in constraint-satisfaction models that might seem feasible is to use the output-activation of the explanation nodes of one run as the input activation of these nodes in the next run (compare left side of Figure 1). Thus, explanation nodes are not reset after each run, but they start with the activation they obtained in the last run. The observation of symptoms is modeled by connecting the currently observed symptom to the SEU. This model is referred to as Simulation 1 in the following.

The reason why this method is not working is the continuous influx of activation from the SEU through the currently observed symptom. Any activation at the beginning of a run is overwritten by spreading activation and only the connection strengths determine the stable state of the network. This can be easily demonstrated by comparing the results of Simulation 1 with a model that is identical despite the fact that the explanation nodes are reset to zero after each run (Simulation 2). Simulation 1 and Simulation 2 produce exactly the same activation results. Both do not capture the change in memory activation.

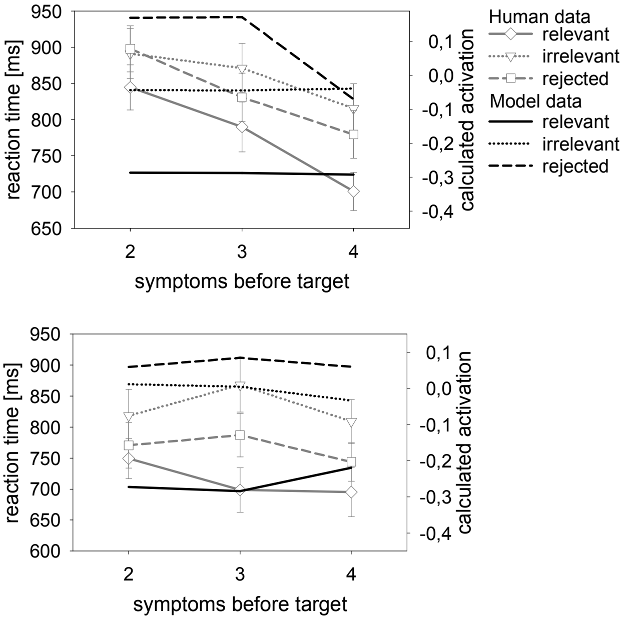

Figure 5. Activation calculated in the constraint satisfaction models (Simulation 1 and 2) and reaction times obtained in the human experiments for consistent (top) and inconsistent trials (bottom). (Activation values are inverted so that they can be plotted directly against the reaction time data.)

In Figure 5, the activation-values calculated by these models are plotted against the human data for consistent (r = -.58) and inconsistent trials (r = -.63). As shown by the graphs and the low correlations between human and model data, the models have an overall bad fit. Although relevant explanations are activated highest in the models as well as in the human data, the increasing activation of these explanations during the course of the trials is not fit by the models. In inconsistent trials, the model-activation even decreases with an increasing number of observed symptoms. Furthermore, contrary to human data, rejected explanations in the model are activated less than irrelevant explanations. Such a pattern of activation should only be expected if incoming information is not integrated properly.

Connection strength to the SEU

Simulation 3 An alternative approach to model sequential data in constraint-satisfaction models is to use the connection strength between the evidence nodes and the SEU as proposed in UECHO (compare Figure 4 and right side of Figure 1). Contrary to Simulations 1 and 2, not only the current symptom but all symptoms observed so far are connected to the SEU. The strength of links to the SEU varies depending on the time elapsed since the respective symptom was observed. The most recently observed symptom gets a full connection to the SEU (.1). Earlier observations are connected to the SEU with a decayed strength. Before each run, the network is reset to its default values. That is, the activation of all chemicals and of all but the currently observed symptoms is set to zero.

Again, one model was run for the consistent (r = -.66) and one for the inconsistent trials (r = -.74). As illustrated by Figure 6, these models produced a much better fit than Simulations 1 and 2. As in the human data, relevant explanations receive the highest and irrelevant explanations receive the lowest activation. However, the models again fail to produce the increasing activation of explanations over the time course.

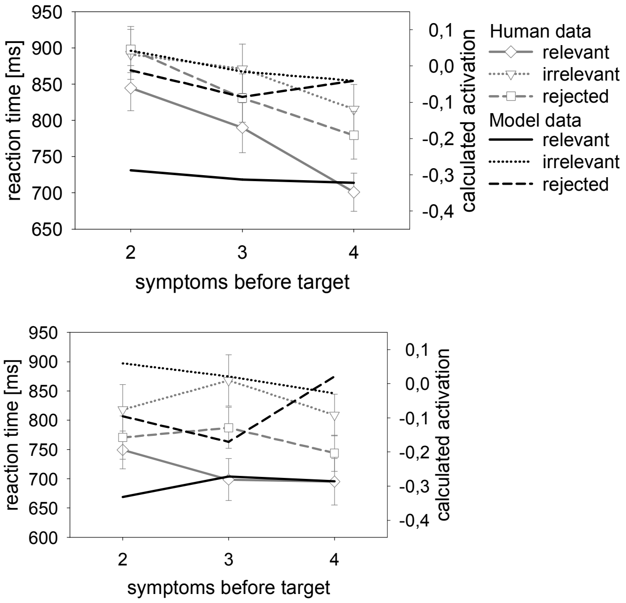

Figure 6. Activation (simulation 3) and reaction times for consistent (top) and inconsistent trials (bottom).

Simulation 4 For better capturing the increasing activation over time, we presumed the influence of each single symptom would need to be higher. Therefore, we developed a fourth set of models with a higher weight given to the full connection between observed symptoms and the SEU. It was not set to .1, as proposed by Wang et al. (2006), but to 1. Except for this change, the models were identical to the models in Simulation 3.

Results of these models are shown in Figure 7. For consistent trials, the model produces the expected effect. Differences between the explanations are fit better (however, they are even overestimated now). Additionally, the model captures the increase of activation over time. This better fit is confirmed by a slightly higher correlation with the human data for consistent (r = -.70) and for inconsistent trials (r = -.81).

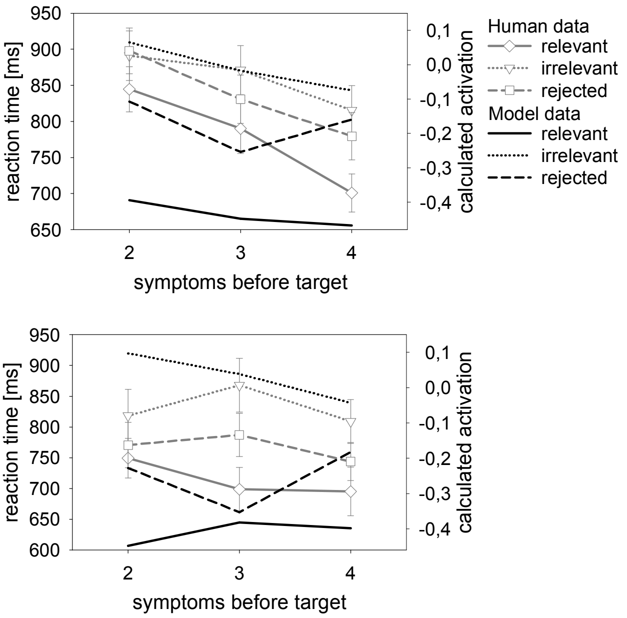

Figure 7. Activation (Simulation 4) and reaction times for consistent (top) and inconsistent trials (bottom).

In Simulations 3 and 4, the previous state of the model is retained by connecting not only the current, but also previous symptoms to the SEU. By letting the strength of these connections decrease over time, the order of observed information is modeled. But is the decay of connection strengths necessary to model sequential information integration?

Simulation 5 To clarify this question, we developed a fifth set of models where, as above, all previously observed symptoms are connected to the SEU. However, previous symptoms do not decay, but they keep the full connection strength of 1.

For consistent trials, this simplified version of the model fits surprisingly well with the human data (r = -.75). The model reproduces the activation differences between the three types of explanations, and the increasing activation over the time course of the task. However, the weakness of the model becomes obvious in inconsistent trials. Whereas the participants' reaction times reflect a change in their diagnosis in the light of the new, inconsistent evidence, the model does not produce a clear difference between relevant and rejected explanations in terms of activation.

Conclusion

We evaluated two possible approaches for modeling sequential information integration in diagnostic reasoning. These approaches differed in the mechanism implemented to integrate new information with information obtained earlier. In the first approach, results from the previous integration step were carried over to be integrated with the new information. In the second approach, the previous state of the network was preserved more indirectly, by connecting not only the current, but also earlier observations to the "energy source" of the network.

Results show that the first approach (Simulations 1 and 2) is not working. No matter what initial activation is used in the network, it is overwritten by the activation resulting from the connection to the SEU. The second approach was more successful. We implemented versions of models that differed with respect to how strongly observed symptoms influenced the current activation of the network (Simulations 3-4). Both models were able to reproduce the activation differences between explanations found in the human data. However, the models differed in their ability to reproduce the courses of explanations' activation over time. A simplified version of these models (Simulation 5), where the influence of earlier evidence did not decay over time, produced a surprisingly high fit in consistent trials, but failed to model the activation data in inconsistent trials.

Concluding, our results support the approach for modeling sequential information integration as it was proposed by Wang et al. (2006). However, our results suggest the parameter setting proposed by Wang et al. to be reconsidered. To adequately model the course of activation during the task, a much higher amount of activation spreading from the observed symptoms needs to be implemented.

We must stress that none of the models was able to sufficiently fit the pattern of activation in inconsistent trials. Although Simulations 3 and 4 produce at least the differences between explanations, they did not model the course of activation during inconsistent trials adequately. This might have several reasons. First, the implementation of constraint satisfaction may be inappropriate. Second, and more plausible given the success of constraint satisfaction models in various areas, the deviation between human and model data demonstrates the involvement of more conscious reasoning processes during inconsistent trials. In consistent trials, the automatic activation processes modeled by the constraint satisfaction networks is perfectly sufficient to solve the task. In inconsistent trials however, a pure activation based approach struggles. Nodes would have to be added or connections other than connections to the SEU would have to be manipulated. To fully capture cognitive processes involved in such trials and in tasks with more complex knowledge structures, hybrid modeling approaches, for example production systems including network dynamics such as ACT-R, might be promising.

Acknowledgements

We thank Matthias Henning, Anja Greiner, Tibor Petzoldt, Josef Krems and four anonymous reviewers for their helpful comments on earlier versions of this paper.

References

Arocha, J. F., & Patel, V. L. (1995). Construction-integration theory and clinical reasoning. In C. A. Weaver, III, S. Mannes & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch.* (pp. 359-381). Hillsdale, Lawrence Erlbaum Associates, Inc.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behaviour, 22*, 261-295.

Baumann, M.R.K., Mehlhorn, K., & Bocklisch, F. (2007). The activation of hypotheses during abductive reasoning. *Proceedings of the 29th Annual Cognitive Science Society* (pp. 803-808). Austin, TX: Cognitive Science Society.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.

Glöckner, A., & Betsch, T. (2008). Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgment and Decision Making, 3*(3), 215–228.

Hagmayer, Y., & Waldmann, M. R. (2002). A constraint satisfaction model of causal learning and reasoning. *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 405-410). Mahwah, NJ: Erlbaum.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24,* 1-55.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Mehlhorn, K., Baumann, M., & Bocklisch, F. (2008). Activation or Inhibition? Why Reasoners are Not Blind for Alternative Explanations. *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2083-2088). Austin, TX: Cognitive Science Society

Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227-234.

Thagard, P. (1989) Explanatory Coherence. *Behavioral and Brain Sciences.* 12. pp. 425-502

Thagard, P., & Kunda, Z. (1998). Making sense of people: Coherence mechanism. In S. J. Read and L. C. Miller (Eds.), *Connectionists models of social reasoning and social behavior (pp. 3-26)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Thagard, P. & Shelley, C. (1997) Abductive reasoning: Logic, visual thinking, and coherence. In: M.-L. Dalla Chiara et al (Eds.), *Logic and Scientific methods. Dordrecht: Kluwer,* pp.413-427

Thagard, P. (2000). *Coherence in thought and action.* Cambridge, MA: MIT Press.

Thomas, R. P., Dougherty, M. R., Sprenger, A. M. & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review, 115,* 155-185.

Wang, H., Johnson, T. R., & Zhang, J. (2006). The order effect in human abductive reasoning: an empirical and computational study. *Journal of Experimental & Theoretical Artificial Intelligence. 18(2),* 215–247