Modeling Information Integration with Parallel Constraint Satisfaction

In which I explore different methods of modeling sequential information integration with parallel constraint satisfaction.

An earlier version of this chapter was published as:

Mehlhorn, K. & Jahn, G. (2009). Modeling sequential information integration with parallel constraint satisfaction. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Modeling Information Integration with Parallel Constraint Satisfaction

Chapter

Abstract

An important aspect of human cognition is the sequential integration of observations while striving for a coherent 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 coherent 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: (1) carrying over results from the previous integration step and (2) decaying input strength of older observations. Results of these models for coherent and incoherent sets of observations are compared to human data from a diagnostic reasoning task.

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 et al., 2006b), 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 coherent mental representation in the light of new incoming information. In this chapter, we explore alternative implementations of such processes in connectionist parallel constraint satisfaction models.

Previous research has shown that reasoners hold knowledge structures that reflect the structure of the task in the environment (Anderson & Schooler, 1991; Gigerenzer et al., 1991). 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., Baumann et al., 2007; Kintsch, 1998; Thomas et al., 2008). To maintain a coherent 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. (2006b) have proposed a connectionist model of sequential information integration based on the idea of explanatory coherence that, probably most prominently, was introduced by Thagard (1989a, 1989b, 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, 1989a).

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 coherent

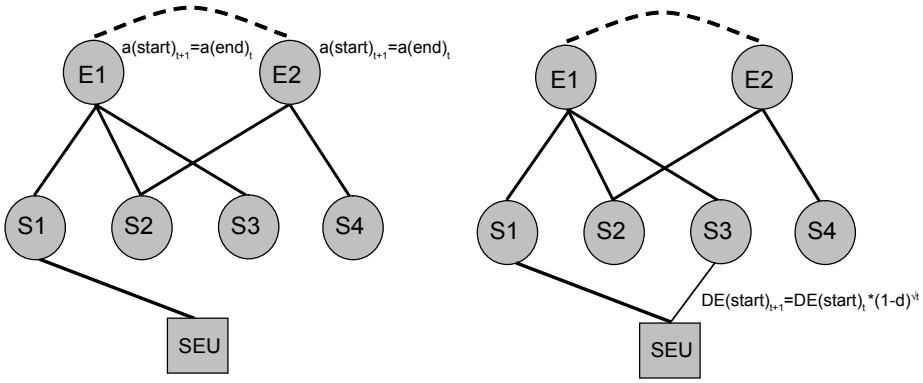

Figure 3.1 Two basic approaches to model sequential data in a constraint-satisfaction network. Either the previous state of the model is preserved by retaining the initial activation of the explanation nodes (left) or previous symptoms keep influencing the activation in the network by a (decaying) connection to the SEU (right).

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, Read, & Miller, 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 in 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; 2006b), 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 up to the

current point in 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 coherently 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 (illustrated in the networks in Figure 3.1). In both networks, the upper nodes, E1 and E2, represent possible explanations of the possible 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 networks, the symptoms S3 and S1 have been observed.

In the left network the previous state of the network is preserved by retaining the activation of the explanation nodes. When the first symptom (S3) is observed, S3 is connected to the SEU and the activation for the explanation nodes (E1 and E2) is calculated. The resulting activation values are used as starting values for the integration of the new symptom (S1).

The right network illustrates the approach proposed by Wang et al. (2006b). Here, the activation of all nodes 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 example, S3 as well as S1 are connected to the SEU. Therewith, the older observation (S3) 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 (S3) are connected to the SEU with a decayed strength. 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. (2006b) 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 in the network. 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: Experiment 1 (Mehlhorn et al., 2008) and Experiment 2 (Baumann et al., 2007). (For a more detailed description of the experiments, see also Chapter 2 of this thesis.) In these experiments, participants diagnosed hypothetical patients after a chemical accident. For each patient, a set of

Table 3.1 Domain Knowledge Participants had to Acquire Before Experiment 1.

Aggregate state and source of contamination	Category	Chemical	Specific symptoms		Unspecific symptoms		
Gasiform, inhaled	Landin	B	Cough	Shortness of breath	Headache		
		T	Cough	Vomiting	Headache		Itching
		W	Cough			Eye inflammation	Itching
Crystalline, skin contact	Amid	Q	Skin irritation	Redness	Headache		
		M	Skin irritation	Shortness of breath	Headache		Itching
		G	Skin irritation			Eye inflammation	Itching
Liquid, drinking water	Fenton	K	Diarrhea	Vomiting	Headache		
		H	Diarrhea	Redness	Headache		Itching
		P	Diarrhea			Eye inflammation	Itching

Note. Original materials were presented in German.

symptoms was presented sequentially on a computer screen and the task was to find the chemical that best explained the set of symptoms. The knowledge necessary to solve this task was taught to participants in an extensive training session. In both experiments, the knowledge consisted of nine different chemicals (named with single letters), grouped into three categories. Each chemical caused three to four symptoms. Symptoms were ambiguous, because each symptom could be caused by two to six different chemicals. Consequently, only the combination of symptoms allowed for unambiguously identifying the correct diagnosis (see Table 3.1 for the knowledge used in Experiment 1).

Two types of trials were used: In both experiments, coherent trials were presented. Additionally, in Experiment 1, incoherent trials were presented (see Figure 3.2 for a coherent and an incoherent sample trial). In coherent trials, all symptoms coherently pointed toward one explanation. Thus, the participants' initial explanation was supported by all later symptoms. In incoherent trials, the explanation suggested by the first two symptoms was incoherent with the later symptoms. Here, participants needed to revise their initial explanation after observing the third symptom. In such incoherent trials, it should be particularly difficult to integrate symptoms while maintaining a coherent mental representation. In total, in Experiment 1, participants were presented with 384 trials, of which 75% were coherent and 25% were incoherent. In Experiment 2, 340 trials were presented, which were all coherent.

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

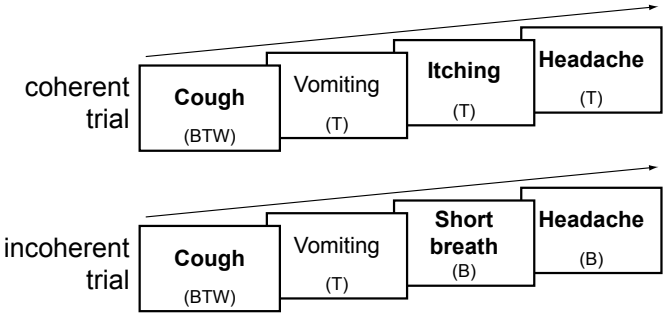

Figure 3.2 Example for a coherent and an incoherent trial in Experiment 1. Letters in parentheses show the compatible explanations after each symptom; they where not visible to the participants.

explanations during the sequential task. This measure is based on the idea of lexical decision tasks (Meyer & Schvaneveldt, 1971) according to which participants should respond faster to a probe that is highly 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 (targets) or one of nine other letters (foils). 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 probe was presented in each trial. Using this measure, it was possible to monitor the activation of explanations over 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. (2006b) asked their participants for explicit belief ratings after each new observation. However, explicitly asking participants during the course of the task might influence the outcome of the task itself (Hogarth & Einhorn, 1992). Directly assessing the activation in memory with an implicit task reduces such a possible influence.

In this chapter, we use response times to target-probes (chemical names) for three different types of explanations to test the constraint satisfaction models. First, we are interested in explanations that are compatible with all symptoms observed before the probe's presentation (*compatible explanations*). Second, we are interested in explanations that are compatible with the initial symptoms, but that are incompatible with later symptoms (*rejected explanations*). Third, we look at explanations that are incompatible with at least the first symptom of the trial (*incompatible explanations*). Reactions to probes for the three kinds of explanations are compared at three different times of measurement over the course of the reasoning task (after *two*, *three*, and *four* symptoms). In Experiment 1 rejected explanations were only presented in incoherent trials. Therefore, below we report the data from the incoherent trials of Experiment 1 and compare them to the coherent trials from Experiment 2.

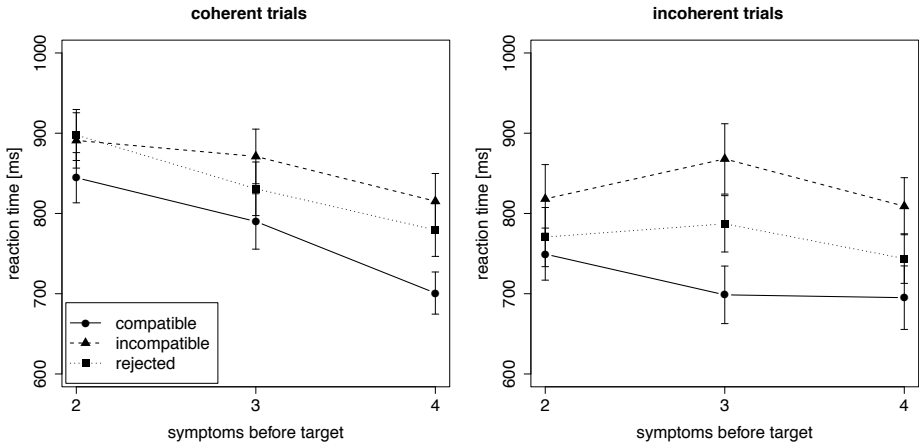

Figure 3.3 Reaction time to compatible, rejected, and incompatible target probes after the 2nd, 3rd, and 4th symptom. Left: coherent trials from Experiment 2 (Baumann et al., 2007); Right: incoherent trials from Experiment 1 (Mehlhorn et al., 2008).

Results

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

Probe reaction task. In both types of trials, the fastest probe responses occurred for compatible explanations. Rejected explanations were responded to slower than compatible explanations, but faster than incompatible explanations (see Figure 3.3; see also Chapter 2 of this thesis for a more elaborate analysis of the behavioural data from the experiments). Coherent and incoherent trials differed in the course of activation over time. In coherent trials, reaction times decreased with an increasing number of symptoms, with the highest decrease for compatible explanations. In incoherent trials, this decrease was less visible, possibly because integrating the information was more difficult than in coherent trials. Nevertheless, the fast responses to compatible explanations suggest that, also in incoherent trials, participants managed to integrate the symptoms correctly.

Models

To assess the validity of the alternative modeling approaches, we implemented the knowledge used in the experiments into different constraint-satisfaction networks (see Figure 3.4 for an example). All networks consisted of the complete material participants needed to learn before the experiment. We used 9 nodes representing

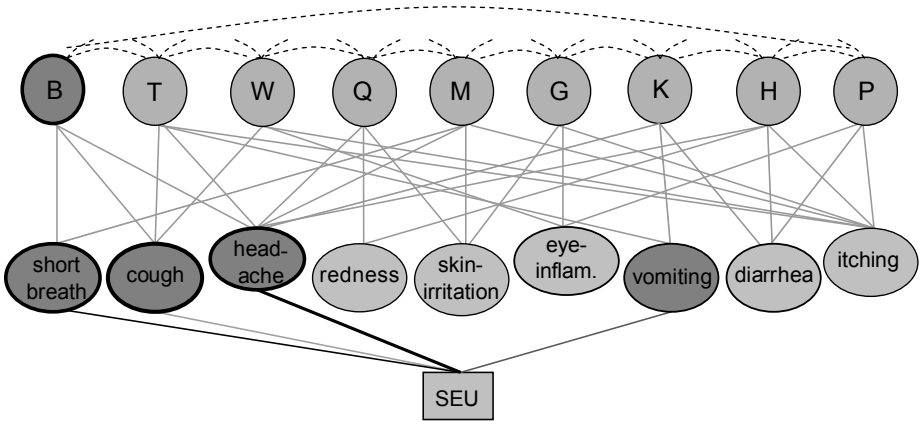

Figure 3.4 Network for an incoherent trial (cough - vomiting - short breath - headache) in Simulation 3. Dashed lines: inhibitory connections, solid lines: excitatory connections. B has the strongest activation when the network settles.

the symptoms, 9 nodes representing the explanations (chemicals), and 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. Symptoms were connected to their associated explanations by excitatory links.

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 values of parameters 2 and 3 were set to fixed values: The initial activation of symptom nodes (parameter 2) was set to 1 for the currently observed symptom and to 0 for all other symptoms. The connection-strength between the nodes in the network (parameter 3) was 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 (compatible, rejected, and incompatible) at the three different times of measurement (after two, three, or four symptoms). This activation is compared to the human probe reaction-time data, which indicates memory activation of explanations.

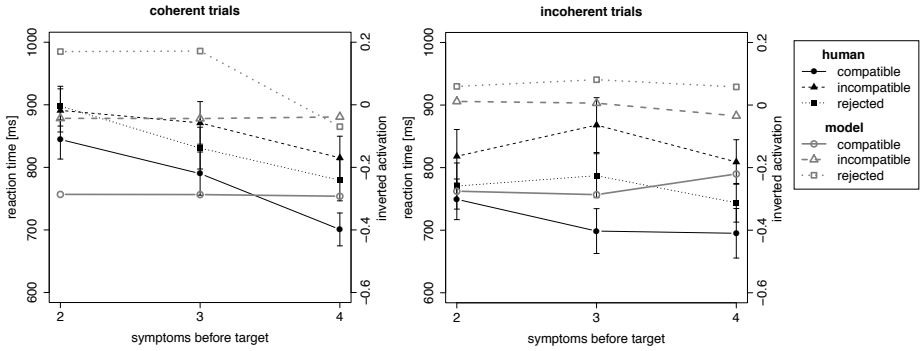

Figure 3.5 Inverted activation values from Simulation 1 and human reaction times for coherent (left) and incoherent trials (right). (Activation values are inverted so that they can be plotted directly against the reaction time data.)

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 (left side of Figure 3.1). Thus, activation values of explanation nodes are not reset before integrating a new symptom, but the values that resulted when integrating the previous symptom are used as start values. The observation of symptoms is modeled by connecting the currently observed symptom to the SEU (with a connection strength of .1). Subsequently, this model is referred to as Simulation 1.

The reason why this method does not work 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 the activation spreading from the SEU and only the connection strengths to the SEU determine the stable state of the network. This can be easily demonstrated by comparing the results of Simulation 1 to a model that is identical except for the fact that the explanation nodes are reset to zero after each run (Simulation 2). Simulation 1 and Simulation 2 produce basically the same activation results.

In Figure 3.5, the inverted activation values calculated by Simulation 1 are plotted against the human data for coherent (r = -.58) and incoherent trials (r = -.63). The model has an overall bad fit. Although compatible explanations are activated highest in the model as well as in the human data, the model does not show an increasing activation over the course of the trials as it is found in the human data. In incoherent trials, the models' activation even decreases with an increasing number of observed symptoms. Furthermore, contrary to the human data, rejected explanations in the model are activated less than incompatible explanations. Such a pattern of activation should only be expected if incoming information is not integrated properly.

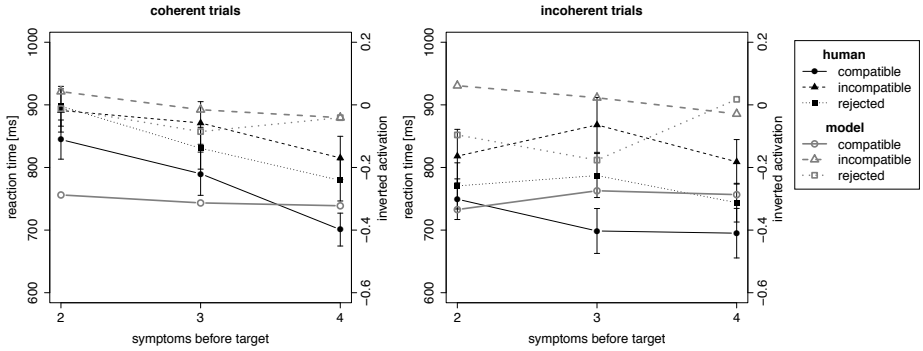

Figure 3.6 Inverted activation values from Simulation 3 and human reaction times for coherent (left) and incoherent trials (right).

Connection Strength to the SEU

Simulation 3. An alternative approach for modeling sequential data in constraint-satisfaction models is to use the connection strength between the evidence nodes and the SEU as proposed by Wang et al. (2006b) (see Figure 3.4 and right side of Figure 3.1). Contrary to Simulations 1 and 2, here not only the current symptom but also previously observed symptoms are connected to the SEU. The strength of the links to the SEU depends 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 as proposed by Wang et al. 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 symptom is set to zero.

Again, the model was run for coherent (r = -.66) and for incoherent trials (r = -.73). As illustrated by Figure 3.6, this model produced a better fit than Simulations 1 and 2. As in the human data, compatible explanations receive the highest activation and incompatible explanations receive the lowest activation. However, the model again has difficulties to fit the change in activation over time. For example, in coherent trials, the model strongly underpredicts the increasing activation of compatible explanations over time.

Simulation 4. For better capturing the increasing activation over time, we presumed that the influence of each single symptom would need to be higher. Therefore, we developed a fourth model in which higher weights were given to the connection between observed symptoms and the SEU. The full connection, that is, the weighting of the most recent symptom, was now set to 1, and the respective decayed connections were calculated based on this value for the full connection. Except for this change, the model was identical to Simulation 3.

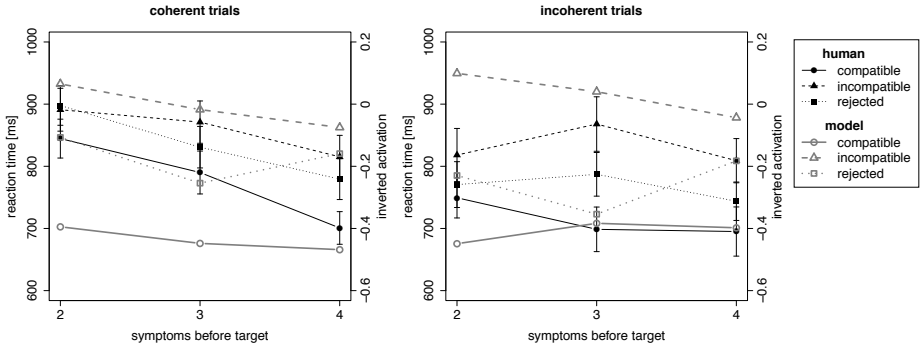

Figure 3.7 Inverted activation values from Simulation 4 and human reaction times for coherent (left) and incoherent trials (right).

Results of this model are shown in Figure 3.7 for coherent (r = -.70) and incoherent trials (r = -.81). As was to be expected, the differences between the different explanations increase compared to Simulation 3. Also the course of activation over time is fit better by this model. However, in coherent trials, the model still underpredicts the increase of activation over time for compatible explanations and it produces a pattern for rejected explanations that is not found in the human data. For incoherent trials, overall, the model predicts the difference between explanations, but it does not fit the change of activation over time.

In Simulations 3 and 4, the previous state of the models is retained by connecting not only the current, but also previous symptoms to the SEU. By letting the strength of these connections decay 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 model where, as in Simulations 3 and 4, all previously observed symptoms are connected to the SEU. However, previous symptoms do not decay, but they keep the full connection strength of 1.

Results are shown in Figure 3.8 for coherent (r = -.75) and incoherent trials (r = -.85). For coherent trials, this simplified version of the model produces the best fit to the human data. It shows the activation differences between the explanations and it fits the activation pattern over time considerably well. However, also this model has difficulties in fitting the incoherent trials. Whereas the participants' reaction times reflect a change in their diagnosis in the light of the new, incoherent evidence (the third symptom of the trial), the model does not produce a clear activation difference between compatible and rejected explanations after the incoherent evidence is observed.

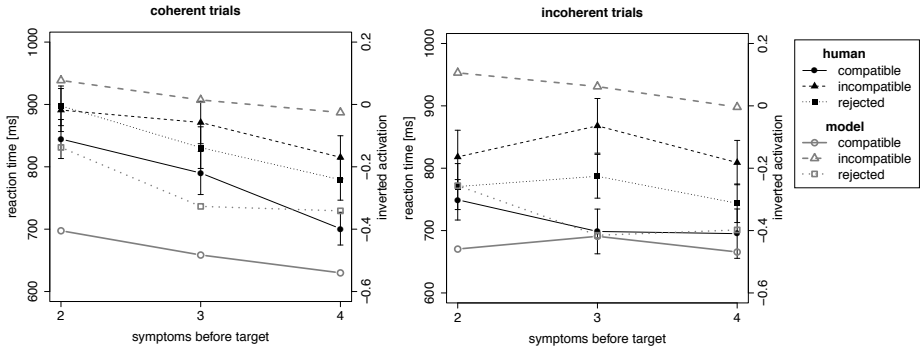

Figure 3.8 Inverted activation values from Simulation 5 and human reaction times for coherent (left) and incoherent trials (right).

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, activation results from the previous integration step were carried over to the next step, where they were 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" (the SEU) of the network.

Results show that the first approach (Simulations 1 and 2) does not work. The initial activation of the network's nodes is overwritten by the activation spreading from the SEU. The second approach was more successful. Following a suggestion from Wang et al. (2006b), we implemented versions of models that differed with respect to how strongly symptoms that were observed over time influence the current activation of the network (Simulations 3 and 4). Both models were able to reproduce the activation differences between explanations found in the human data, with higher weights of the connection between SEU and observed symptoms resulting in better model fits. However, also these models had difficulties in fitting the course of 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 coherent trials, but failed to model the time course of activation in incoherent trials.

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

that not yet explained observations do not decay over time, as suggested by Wang et al., but retain a stable influence on the network.

We must stress that none of the models was able to sufficiently fit the pattern of activation in incoherent trials. Although Simulations 3 and 4, where observations decayed over time, produced at least the differences between explanations as found in the human data, they did not model the course of activation 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 incoherent trials. In coherent trials, the automatic activation processes modeled by the constraint satisfaction networks is perfectly sufficient to solve the task. In incoherent 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. As we discussed in more detail in Chapter 2, to fully capture cognitive processes involved in such trials and in tasks with more complex knowledge structures, hybrid modeling approaches which allow for investigating the interaction of automatic memory activation with more deliberate reasoning strategies, might be promising.