Introduction

In which I give an overview of the topic and introduce the methods.

Chapter 1

Introduction

Introduction

It was at the beginning of my graduate time, when, during a guest lecture at the University of Chemnitz, Frank Ritter talked about his favorite scientific articles. One of them was Allen Newell's 20 questions paper (1973). Newell had argued that psychology focuses too much on isolated, experimental phenomena and simplifying dichotomies, rather than working towards a precise and unified theory of cognition. If I would have read this paper in more detail then, and truly understood what Newell meant, working on my dissertation might have gone more smoothly. But I did not do that. Rather, I began working in "good psychological tradition". I had studied my theories, I knew how to set up experiments and do an ANOVA, and I thought that was sufficient to investigate cognition.

The starting point of my dissertation was the idea that automatic memory processes are an important aspect underlying decision making. Specifically I was interested in the role of memory activation in diagnostic reasoning. Diagnostic reasoning is the reasoning from observed data to explanations and involves the generation and evaluation of hypotheses that represent potential explanations. I wanted to know why, when confronted with a number of medical symptoms, possible diagnoses seem to pop up almost effortlessly in a physician's head. And, why, when being in a certain context, one cannot help but interpret new information in the light of this context. My idea was that these phenomena were largely due to automatic memory processes, which make information that is associated to the current context (e.g., observed medical symptoms) available in memory. Such available information could then be subjected to more deliberate reasoning processes as they had been classically discussed in the reasoning literature. While the idea of automatic activation processes regulating the availability of memory contents was not new, direct experimental evidence for such memory processes in diagnostic reasoning was sparse.

With the goal to present such evidence, we set out to conduct a series of experiments (Baumann, Mehlhorn, & Bocklisch, 2007; Mehlhorn, Baumann, & Bocklisch, 2008). In these experiments, we used a probe reaction task to track the availability of different diagnostic hypotheses in memory, while participants had to generate diagnoses for sequentially presented medical symptoms. The probe reaction task was based on the idea of lexical decision tasks, where participants respond faster to a probe that is more highly activated in memory than to a probe of lower activation (e.g., Meyer & Schvaneveldt, 1971). If observations indeed activate associated explanations in memory, then, when presented with symptoms like fever, nausea, and headache, a participant should react faster to the probe "influenza", than to a probe that is less related to these symptoms (e.g., "pregnancy"), or to a neutral probe (e.g., "house"). To avoid the possible influence of previous experience on memory activation, in the experiments we used artificial medical knowledge, which consisted of medical symptoms that were caused by hypothetical chemicals. Chemicals were named with single letters, which allowed us to use letters in the probe reaction task, thereby preventing potential problems associated with the use of complete words (e.g., individual differences in reading speed and word frequency effects).

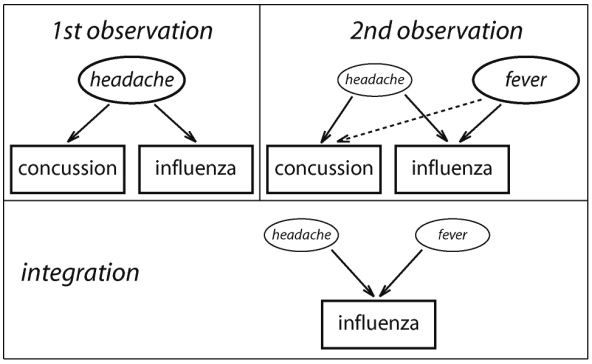

Figure 1.1 Box and arrow model of memory processes assumed to underlie diagnostic reasoning as reported in Baumann et al. (2007).

Overall, the results seemed to support our theoretical considerations. For example, diagnostic hypotheses that were compatible to all observed symptoms caused the fastest probe reactions, suggesting that these compatible hypotheses were indeed more easily available in memory than their alternatives. Also, with an increasing amount of observed symptoms, reaction times to compatible hypotheses decreased faster than reaction times to other hypotheses. This suggested that the availability of hypotheses indeed might be a function of their association to observed symptoms. However, when trying to understand the results in detail, we quickly ran into open questions. Could it have been that the results were actually not caused by memory activation, but were merely byproducts of deliberate reasoning processes? And, assuming that it was indeed memory activation that caused our results, how would the underlying activation processes look precisely? What we needed was a detailed model of the assumed memory processes that would make precise predictions.

The first "model" that we generated consisted of several boxes and arrows (see Figure 1.1 and Baumann et al., 2007). The boxes represented medical symptoms (e.g., headache) and their potential explanations (e.g., influenza). The arrows represented the associations between symptoms and explanations that could be positive (solid lines) or negative (dashed lines). This model was useful to illustrate the processes that we assumed to cause the activation of diagnostic hypotheses. For example, we proposed that "After the integration phase the influenza explanation as a still relevant explanation should be strengthened as it receives activation both from the symptom fever and the symptom headache [...]" (Baumann et al., 2007, p. 804). However, the problem with this model was a lack of precision. For example, why exactly was there a positive association between headache and influenza and how strong was it precisely? As Allen Newell (1973) put it, in such a model "Too much is left unspecific and unconstrained." (p. 301).

The lack of theoretical precision, which we faced when trying to understand our data, is inherent to verbal theories of cognition (and their associated box and arrow models). A solution for that problem has been proposed by Allen Newell (1990) and many others. It is the use of computational cognitive models. These models should be specific and constrained enough to provide quantitative predictions that can be tested by comparing them to human data. After the initial difficulties described above, I moved on to using such models. In the remainder of the introduction I give a brief introduction of the two modeling approaches that I used in my dissertation and present an overview of the chapters in this thesis.

A Connectionist Approach: ECHO

In his *theory of explanatory coherence*, Thagard (1989a, 1989b, 2000) proposes that a set of propositions (e.g., observations and their potential explanations in memory) can be evaluated by automatic activation processes, purely on the basis of their coherence. In the connectionist constraint-satisfaction implementation of this theory, ECHO (e.g., Thagard, 1989a), propositions are represented by a network of interconnected nodes. The connections between the nodes represent the relations (constraints) between the respective propositions. Depending on these connections, when the network is integrated, activation or inhibition is spread between the nodes. After the network has been integrated, the strength of a proposition is indicated by the numerical activation of its node, which depends on its coherence to the other nodes in the network. Applying Thagard's theory to diagnostic reasoning predicts that those explanations that are strongly associated with the observed data are most strongly available in memory (because they receive a large amount of activation) and that less strongly associated explanations have a lower availability (because they receive less activation and potentially also inhibition).

As we will show in Chapter 3, such a connectionist account increases the precision compared to mere verbal predictions. It requires a detailed specification of the assumed memory processes (e.g., how strong is the connection between observation x and explanation y?) and it predicts precise numerical activation values that can be compared to behavioral data. However, this account has also some major limitations (see e.g., Fodor & Pylyshyn, 1988, for an overview). Maybe most importantly, it does not represent a fully functioning cognitive system. While presenting a precise account of activation dynamics within an assumed network, it remains mute about the interplay of these dynamics with, for example, perceptual, decisional, intentional, and motor processes, which might play an important role in human reasoners. Another problematic point is the interpretability of its results. The model predicts precise activation values, which can be plotted against and correlated with behavioral data. But what exactly do these values mean and how, precisely, do they correspond with behavioral data?

An Architectural Approach: ACT-R

An approach that not only endeavors precision, but also comprehensiveness in terms of understanding how the brain "achieves the function of the mind" (Anderson, 2007, p. 7) is the use of cognitive architectures[1]. The term "cognitive architecture" was, maybe most prominently, described by Allen Newell (1990) as a way towards the "ultimate goal" of a unified theory of human cognition. The idea is that the architecture is both a psychological theory, as well as a platform for constructing computational models, that allows for investigating different phenomena within one framework. This idea has been developed since then, resulting in various architectures, like for example EPIC (Meyer & Kieras, 1997), Soar (A. Newell, 1990), and ACT-R (Anderson et al., 2004).

The cognitive architecture I used in my dissertation is ACT-R, because it puts a strong emphasis on processes underlying memory activation. It has received empirical support and validation from a large number of studies in a variety of research areas, ranging from list memory (Anderson, Bothell, Lebiere, & Matessa, 1998) to car driving (Salvucci, 2006). ACT-R allows for modeling of the complete task as solved by the participant. Thereby, without requiring additional assumptions about how the model maps on the experiment, it produces results that are directly comparable to human data. This is possible because the underlying theory makes precise predictions not only about the probability and latency of retrieving facts from memory, but also about the time needed to perceive stimuli and give responses.

In ACT-R, cognition is described by a number of independent modules. Each of the modules represents a different cognitive resource and is associated with specific brain regions. For example, a visual module allows ACT-R to perceive visual stimuli and a motor module allows for motor actions like pressing a key. Most important for the work presented in this thesis are three of the central cognitive modules: the imaginal module, the declarative module, and the procedural module.

The imaginal module holds information necessary to perform the current task and is thereby comparable to the focus of attention in working memory (e.g., Borst, Taatgen, & van Rijn, 2010). In a diagnostic reasoning task, the imaginal module might, for example, hold observed medical symptoms, which determine the present usefulness of potential explanations. In Chapter 2 we investigate how such observed symptoms can affect the availability of explanations in long-term memory.

The declarative module allows for the storage in and retrieval of facts from declarative memory and thereby represents ACT-R's account of long-term memory. In a diagnostic reasoning task, such facts could, for example, be possible diagnoses. Availability of the facts is determined by their activation (see Chapters 2, 4, and 5 for a detailed description of the underlying equations). Basically, the activation of a fact

[1] *In the literature, the term cognitive architecture has also been used for connectionist models (e.g., Kintsch, 1998). In this thesis we use the term cognitive architecture exclusively for what Fodor and Pylyshyn (1988) referred to as "Classical architectures", that is, architectures that are committed to a symbol-level of representation and thereby aspire "paying attention to three things: the brain, the mind (functional cognition), and the architectural abstractions that link them" (Anderson, 2007, p. 8). However, as Fodor and Pylyshyn (1988) point out, connectionism might provide "an account of the neural [...] structures in which Classical cognitive architecture is implemented" (p. 3). In fact, Lebiere and Anderson (1993) successfully created such a connectionist implementation of an early version of ACT-R.*

represents the likelihood that it will be needed in the near future and depends on two factors: its past and present usefulness. In Chapter 4 we explore the respective contribution of these two factors for the availability of hypotheses in diagnostic reasoning.

The procedural module allows for communication between the other modules. It contains production rules, which can recognize patterns of information in the modules' so-called buffers, and react to these patterns by sending requests to the modules. A production rule might, for example, recognize that a visually presented symptom was encoded in the visual buffer and react by requesting the name of this symptom from declarative memory. Production rules implement strategies that the reasoner might use in a certain situation. For example, after retrieving an explanation for observed medical symptoms from memory, one strategy might be to simply give that explanation as diagnosis, whereas another strategy would be to deliberately test the explanation against potential alternatives. In Chapter 5 we use this module to implement different decision making strategies and test how well these strategies predict behavioral data.

Overview

In this thesis I will show how we used the approaches outlined above to implement and test precise models of decision making.

In Chapter 2, we introduce our idea of how memory activation affects the availability of explanations. We present several ACT-R models that all share the assumption that observations stored in working memory can activate associated explanations in long-term memory. The models differ in their assumptions about how sequentially observed symptoms affect the activation of associated explanations over time. Using ACT-R allows us for testing these assumptions within a well-established and elaborate theory of human memory. It also allows for investigating the interaction of the assumed memory processes with other potentially task-relevant factors. The results of the models are compared to human data from two behavioral experiments in which we used the probe reaction task mentioned above to track the availability of different explanations during a sequential diagnostic reasoning task.

In Chapter 3, we explore different methods of modeling sequential information integration with connectionist constraint satisfaction models, based on Thagard's ECHO. Just like the ACT-R models presented in Chapter 2, the models share the basic assumption that observations can activate associated explanations, but they differ in how sequentially observed medical symptoms affect the activation of explanations over time. The models are evaluated on the probe reaction data from the same experiments as presented in Chapter 2.

In Chapter 4, we investigate how an explanation's present usefulness, as reflected by the observed symptoms, interacts with its past usefulness, as reflected by the recency and frequency of previous encounters with the explanation. We thereby test whether the memory mechanisms as proposed by the ACT-R theory can explain why, out of all possible hypotheses, reasoners tend to generate those hypotheses from memory

that have a high a priori probability and a high usefulness in the current context. Model predictions are compared to behavioral data from an experiment in which we manipulated both memory components independently, by means of a secondary task that had to be solved next to a primary diagnostic reasoning task.

In Chapter 5, we move on to a slightly different domain of decision making. Whereas in Chapters 2 to 4 we investigate how automatic activation process affect the availability of information in memory as a function of the past and present environment, in Chapter 5 we investigate how reasoners use information from memory, given its availability. More specifically, we focus on a debate that has evolved over the last decade in the decision-making literature and is centered on the question whether decisions can better be described by simple non-compensatory heuristics or by more complex compensatory decision making strategies. In Chapter 5 we show how the precision and comprehensiveness provided by a cognitive architecture can be used to get beyond the simple dichotomy of non-compensatory versus compensatory decision strategies. We use ACT-R to implement various strategies that have been discussed in the literature and compare the resulting quantitative predictions to behavioral data from two previously published experiments (Pachur, Bröder, & Marewski, 2008).