The Influence of Causal Interpretation on Memory for System States

Wolfgang Schoppek (wolfgang.schoppek@uni-bayreuth.de)
Department of Psychology, University of Bayreuth
D-95440 Bayreuth, Germany

Abstract

This paper reports an experiment that investigated the influence of causal interpretation on acquisition and use of two knowledge types about a static system: I-O knowledge (instances of system states) and structural knowledge (knowledge about causal relations within the system). One group of subjects saw system states without being informed about the causal nature of the material. Another group saw the same states as switches and lamps. It is assumed that the group without causal interpretation can only acquire I-O knowledge. If I-O knowledge is the predominant type when dealing with small systems, then there should be no group differences in a recognition task. Actually, the group with causal interpretation discriminates much better between targets and distractors, but with longer RTs. This is interpreted in terms of structural knowledge acquired by the group with causal interpretation, which was used to reconstruct system states in cases of doubt. Results of a task where subjects had to judge single causal relations support that interpretation, but also indicate that the knowledge about effects is probably not represented in an explicit, symbolic form. An ACT-R model that uses associations between events as a subsymbolic form of structural knowledge reproduces the data well. Thus, data and model support the significance of I-O knowledge but also shed some light on the role and the development of structural knowledge.

One central question in the psychological research on complex dynamic systems refers to the knowledge that is used for controlling a system. One important aspect of that question refers to the content of the acquired knowledge. Subjects may acquire structural knowledge, defined as general knowledge about the variables of a system and their causal relations. They may as well acquire input-output knowledge (I-O knowledge), which represents instances of input values and the corresponding output values.

There is evidence for the influence of both types of knowledge on performance in system control, but currently many authors emphasize the role of I-O knowledge, particularly when dealing with small systems like the "Sugar Factory" (a dynamic system with one input and one output variable, connected by a linear equation; Berry & Broadbent, 1988). Computational models developed on the basis of Logan's Instance Theory (Dienes & Fahey, 1995) or ACT-R (Lebiere, Wallach & Taatgen, 1998) demonstrate the sufficiency of I-O knowledge for the control of the "Sugar Factory". The strategy of relying on I-O knowledge seems to be preferred by most subjects, even in the control of more complex systems. However, in systems of at least six variables, high performance is usually associated with structural knowledge (Funke, 1993; Vollmeyer, Burns & Holyoak, 1995).

A second aspect of the question as to what knowledge is used in system control refers to its status as explicit or implicit knowledge. In an experiment with the "Sugar Factory", Dienes and Fahey (1998) found stochastic independence between the solution of studied control problems and the recognition of the same situations as studied. The authors concluded that memory for the situations was implicit. This result extends the common finding of dissociations between recognition and completion tasks (e.g. Tulving & Hayman, 1993) to the domain of system control.

In the present paper these questions were investigated by using stimuli that can be either interpreted as states of a system or simply as spatial patterns. The rationale of the experiment is that learning of instances does not depend on the causal interpretation of stimuli. Consequently, if knowledge about instances (I-O knowledge) is the main knowledge type learned, there should be no effect of causal interpretation on recognition of system states. On the other hand, if structural knowledge is learned additionally, then causal interpretation should have positive effects, particularly in a causal judgment task.

The assumptions about the two knowledge types are explicated with a computational model based on the ACT-R theory (Anderson & Lebiere, 1998). The model reproduces the results of the experiment quite well, and can be considered being an explanation for the stochastic independence between completion and recognition tasks.

Experiment

The significance of I-O knowledge and structural knowledge was studied with a system consisting of four lamps operated by four switches. Figure 1 shows a screenshot with the effects of the switches mapped (the arrows were not visible for the subjects). Each switch affects one or two lamps. Two of the effects are negative, which means that the corresponding lamp is switched off when the switch is turned on.

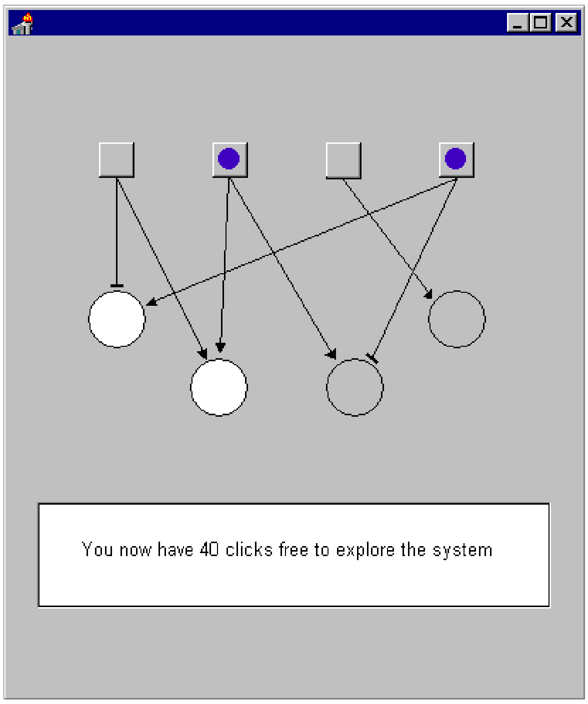

Figure 1: The system used in the experiment.
The arrows were not visible for the subjects;
$\downarrow$: on relation, $\perp$: off relation.

Two tasks were used, each more sensitive to a different type of knowledge: A recognition task - easiest to be done with I-O knowledge, and a causal judgment task - easiest to be done with structural knowledge. Additionally, a pattern completion task was administered, which is not expected to be particularly sensitive to one knowledge type.

In the speeded recognition task subjects saw ten possible and ten impossible system states two times each, and had to decide if they had seen the state in the learning phase or not. The items of the speeded judgment task were pictures of the switches and lamps with one switch and one lamp highlighted. Subjects had to decide if there was a causal relation between the highlighted elements. The 16 possible combinations were shown twice. In the completion task subjects were shown eleven arrays of switches or lamps and asked to complete the missing parts, i.e. complete the lamps when switches were shown and vice versa.

Two factors were varied between subjects: (1) the possibility to interprete the pictures of system states shown in the learning phase as causal, and (2) the subject's activity, i.e. if the system states were either observed, or produced by operating the switches. I will focus on the effects of the first factor (that were the strongest ones, anyway), and report the data of the two groups who observed the system states in the learning phase, either with causal interpretation (ci), or without causal interpretation (nci). Each of the groups consisted of 12 subjects.

Other factors were varied within subjects: (1) the number of presentations of each state in the learning phase (1-2 presentations vs. 3-5 presentations), and (2) the number of switches that were "on" in each item of the recognition task (1 switch on vs. 3-4 switches on).

The experiment started with a learning phase where subjects saw 40 system states in intervals of four seconds. Each possible state of the system was shown at least once. The group without causal interpretation (nci) was told that they would see spatial patterns, which they should memorize. The group with causal interpretation (ci) was informed that the patterns were states of a system of switches and lamps.

Three minutes after completion of the learning phase the recognition task was administered followed by another 25 system states. Next, subjects worked on the completion task. Then the subjects of the group without causal interpretation were debriefed about the causal nature of the stimulus material. After that the judgment task was provided, followed by two other tasks that are not reported here.

Given the assumption that knowledge about the system is primarily stored as specific instances, the factor "causal interpretation" should have no effects on performance in the recognition task. If, however, subjects acquire structural knowledge - which is expected only in the group with causal interpretation - that group should outperform the nci group, particularly in the judgment task.

As a measure of performance in the recognition and judgment tasks, discrimination indices P_r were calculated according to the Two-High-Threshold-Model (Snodgrass & Corvin, 1988). A discrimination index of 1 indicates perfect discrimination; a value of 0 indicates random performance.

Table 1: Discrimination indices for two tasks

	ci	nci
Recognition	M=0.48 s=0.23	M=0.30 s=0.22
causal judgment	M=0.55 s=0.18	M=0.17 s=0.23

Table 1 shows means and standard deviations of these indices. In both tasks the group with causal interpretation is significantly better ($F_{1,22} = 10.76$, $p < .01$), and there is an interaction between task and group ($F_{1,22} = 7.26$, $p < .05$). The ci group is better at judging causal relations than at recognition; for the ci group the reverse is true. Latencies for hits are longer in the group with causal interpretation (ci: 2250 ms, nci: 1493 ms). The fact that the variance is also significantly higher in the ci group points to the use of different strategies: If a system state could not be retrieved in the recognition task, subjects of the ci group might have tried to reconstruct the state by using knowledge about the effects of the switches. That would mean that subjects used both, I-O knowledge and structural knowledge.

This interpretation is supported by the effects of the within-subjects factors on recognition performance (Figure 2, left panel). If the reconstruction hypothesis is true, then there should be an effect of the number of switches on in the ci group, because the reconstruction process is harder the more switches have to be considered. Actually, a significant interaction between group and number of switches on was found in the proportion of hits ($F_{1,22} = 6.13$, $p < .05$). States with three or four switches in on position are particularly badly recognized by the subjects of the ci group. On the other hand, in the group without causal interpretation the influence of number of presentations is higher (interaction marginally significant: $F_{1,22} = 3.63$, $p = .07$). All this supports the assumption that the group with causal interpretation used I-O knowledge and structural knowledge in both tasks.

Further inferences about the application of one vs. two knowledge types can be drawn from contingency analyses between the tasks. Since the mapping between the items of the recognition task and the items of the judgment task is ambiguous, I calculated contingencies between recognition and completion.

If there is only one (explicit) knowledge type, items that were completed correctly should also be recognized as studied. For two knowledge types, the contingency prediction is less clear. If each task is solved with different knowledge, stochastic independence between the tasks should be the consequence.

The items of the completion task were entered into contingency tables depending on their solution and their recognition (e.g. Item 1 was solved correctly and not recognized as studied). The entries were summed over all subjects of each condition and over all items. Empirical contingencies, measured by Δp, were compared with maximum contingencies that can result with the given marginal distributions[1]. In the ci group the empirical contingency between recognition and correct completion is 0.17. This is considerably lower than the maximum of 0.65. In the nci group the contingency is 0.41, which is much closer to the maximum of 0.53 in that group. Thus, in the group with causal interpretation the solution of completion items does not depend on correct recognition of these items as studied, whereas in the group without causal interpretation a moderate degree of dependency was found between the two tasks. Again, the results are compatible with the assumption that the nci group used only one type of knowledge, whereas the ci group used two types.

Discussion

Overall, the results support the assumption that causal interpretation enabled subjects to gain an additional type of knowledge. This raises the question about the nature of that knowledge. In the introduction I hypothesized that it should be structural knowledge. But there is one result that is problematic for this conclusion: Since structural knowledge is ideal for solving the judgment task it is surprising that the mean latency for hits is as long as 2234 ms (see also Schoppek, 1998 for similar results). If subjects tried to retrieve structural knowledge right away, the latency should be much shorter. A possible explanation is that most subjects try to use I-O knowledge first and use knowledge about effects only after retrieval of relevant I-O knowledge fails. The reason for that might be that knowledge about causal relations is not represented explicitly in symbolic form, but rather in form of associations between events. In the ACT-R theory (Anderson & Lebiere, 1998), associations between declarative memory elements and their baselevel activations are described as the subsymbolic level of declarative memory. This level is implicit in the sense that it affects symbolic processing (e.g. retrieval) without being directly accessible. In the next section I describe a computational model that uses the distinction between symbolic and subsymbolic level to explain the effects of causal interpretation.

ACT-R Model

In order to test how the above interpretation can reproduce the data, I developed an ACT-R model that simulates the learning phase, the recognition task, and the judgment task. There are two versions of the model. One of them entails additional production rules for modeling causal interpretation. These rules reconstruct a system state when no relevant memory representation of the state can be retrieved. The state is reconstructed on the basis of associations between events.

In the learning phase a new declarative element (called chunk in ACT-R) is created for each system state and pushed on the goal stack. After processing the goal it represents a system state with its slots holding the arrays of switches and lamps. These state chunks are the basic units of I-O knowledge. Also in each cycle, a change-image is created as a subgoal, representing the changes between the previous and the current system state. Most of the change-images are not strong enough to be retrieved later on, but during goal elaboration associative weights are learned between

[1] The maximum possible memory dependence as suggested by Ostergaard (1992) could not be calculated because only studied items were used in the completion task.

switch- and lamp-events (e.g. between the events "Switch A turned on" and "Lamp 1 turns dark"). Afterwards these associations are used to reconstruct system states in the condition with causal interpretation. No structural knowledge is explicitly induced, because otherwise the model would predict much shorter response times in the judgment task.

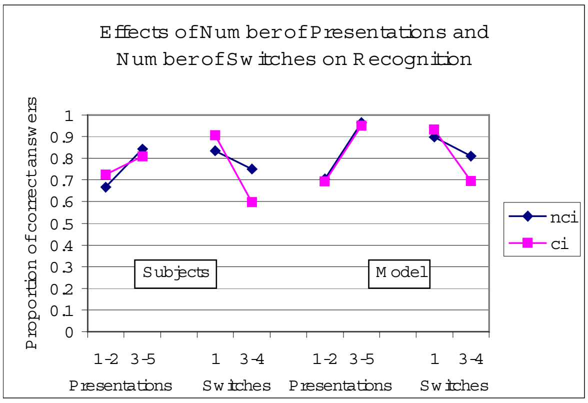

Figure 2: Experimental (left) and model (right) results

In the recognition task both model versions try to retrieve an instance similar to the probe. The constraints for retrieval are either “retrieve a chunk that has the probe’s combination of switches in its switches slot” (retrieval by input), or “retrieve a chunk that has the probe’s combination of lamps in its lamps slot” (retrieval by output). The model has a bias towards using the tactic of retrieval by input. Partial matching is turned on, which means that not only perfectly matching instances can be retrieved, but also instances that are similar to the retrieval constraints. If retrieval fails, the version without causal interpretation guesses, the other version starts the reconstruction process. Reconstruction is based on the lamp-events that are most strongly activated by the switch-events shown in the probe ("switch on"). The probability of false reconstructions rises with the number of switches that are on - an effect that explains the bad recognition performance under condition ci & 3-4 switches.

I simulated two samples with 24 cases each[2]. Some results are shown in the right panel of Figure 2. In both simulated between subject conditions recognition performance depends more on the number of presentations as compared to the real subjects. But the interaction between number of switches and causal interpretation is well reproduced by the model. In general, the model overestimates recognition performance. This effect is mainly due to the excellent recognition of the frequently shown system states. Latencies for hits are very close to the data: 2314 ms in the simulated ci group and 1541 ms in the simulated nci group (note that the latency factor was fitted for the nci group only).

After fitting parameters for the recognition task, the model was extended with a few production rules to solve the causal judgment task. In that task the model tries to retrieve a diagnostic instance appropriate to confirm the causal relation. For example, when the item requires judging the causal relation between Switch A and Lamp 3, the model tries to retrieve a chunk that represents the system state with Switch A as the only switch on. Assume the model retrieves the appropriate state (Switch A on, Lamp 2 on), it will produce the answer “no”. If no diagnostic state can be retrieved, the model reconstructs the state in the same way as in the recognition task.

In the simulation with this part of the model, I assumed that the judgment task was done right after the learning phase. Recall that the groups of subjects that have been discussed so far did the judgment task later in the experiment. Therefore, the simulation results were compared to a group of subjects (N=12) who did the judgment task in the first place. That group was informed about the causal interpretation of the stimuli.

The model matches the subjects' data quite close without fitting any parameters (Figure 3). Mean latencies for hits were 2305 and 2234 ms in the model's and subjects' data, respectively.

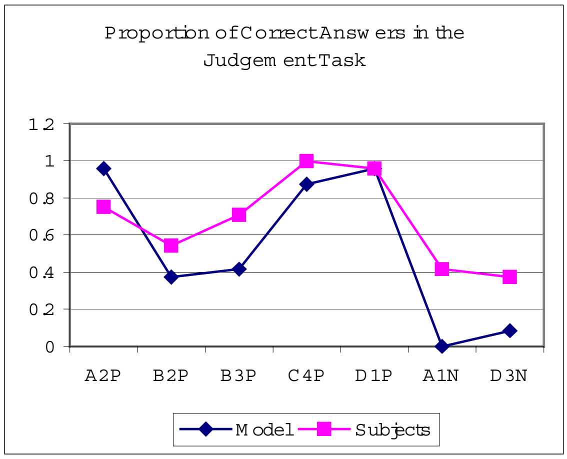

Figure 3: Proportions of correct answers in the judgment task. A2P through D1P are the five "on" relations of the system, A1N and D3N the two "off" relations. (A2P: Switch A – Lamp 2 - positive)

General Discussion

Model and data support the view that I-O knowledge is the primary type of knowledge used when dealing with a small system. But longer latencies, together with

[2] Parameter values were as follows: partial matching=on, mismatch penalty=2.5, baselevel learning=0.5, retrieval threshold=0.75, parameter learning=off, associative learning=3.0, activation noise s=0.5, expected gain noise s=0.5, latency factor=2.5. The source code of the model is available at www.uni-bayreuth.de/departments/psychologie/cogsci01.html

better recognition in the group with causal interpretation point to the use of an additional type of knowledge. It has been modeled as subsymbolic associations between events, used to reconstruct a mental image of the system state in question.

The results of the group with causal interpretation parallel the findings of Dienes & Fahey (1998), but the interpretations are slightly different. Dienes and Fahey assume that subjects learn a lookup-table of system states and conclude from their data that this table is stored in implicit memory. The lookup-table is similar to I-O knowledge. The difference is that in the present conception I-O knowledge is always explicit, and a second type of knowledge is assumed – subsymbolic associations between events. In this interpretation it is the subsymbolic knowledge that would be considered implicit.

Applying the distinction between symbolic I-O knowledge and subsymbolic associations between events to the "Sugar Factory" could explain the results of Dienes & Fahey (1998). If subjects used I-O knowledge about past situations in the recognition task and associations between events in the control problems, stochastic independence between the two tasks could be the consequence. The explanatory potential of the subsymbolic level of ACT-R for implicit memory phenomena has also been demonstrated by Taatgen (1999) with a model of word recognition and completion. In his model it is the dynamics of baselevel learning rather than associative learning that accounts for dissociations.

The present research yielded effects that are similar to those known from other paradigms. It is a common finding that providing additional information about stimuli enhances memory or other kind of performance, e.g. in classification learning (Nosofsky, Clark, & Shin, 1989), Schema acquisition (Ahn, Brewer, & Mooney, 1992), or text comprehension (Bransford & Johnson, 1973; Kintsch & van Dijk, 1978). Also the finding that most subjects spontaneously rather use I-O knowledge or knowledge about specific instances than using structural knowledge or rule knowledge has parallels in these paradigms. Nosofsky et al. (1989) found that even simple rules defining a concept were only used when subjects were explicitly told to do so. Ahn et al.'s (1992) subjects used the experimentally provided background knowledge only when they were engaged in tasks requiring the active use of that knowledge.

An important question is at what point in the whole process the causal interpretation effect arises. The present model assumes that the associations between events are learned incidentally in both conditions, and the effect occurs during recall, when only the ci subjects use this knowledge. This assumption shall be tested in future experiments.

The next step in this research is modeling the completion task to test if the model really predicts the effect of causal interpretation on the contingency between recognition and completion tasks. Further research is also necessary to explore if the effects of causal interpretation can be generalized to similar tasks. If the effects can be confirmed, the model provides an interesting basis for a more general theory about implicit memory phenomena.

References

Ahn, W., Brewer, W. F., & Mooney, R. J. (1992). Schema acquisition from a single example. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 391-412.

Anderson J.R., & Lebiere C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Berry D.C., & Broadbent D.E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, 79, 251-272.

Bransford, J.D. & Johnson, M.K. (1973). Considerations of some problems of comprehension. W.G. Chase (Ed.), *Visual information processing*. Orlando, FL: Academic Press.

Dienes Z., & Fahey R. (1995). Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 848-862.

Dienes Z., & Fahey R. (1998). The role of implicit memory in controlling a dynamic system. *The Quarterly Journal of Experimental Psychology*, 51A, 593-614.

Funke J. (1993). Microworlds based on linear equation systems: a new approach to complex problem solving and experimental results. In G. Strube & K.F. Wender (Eds.), *The cognitive psychology of knowledge*, pp. 313-330. Amsterdam: North-Holland.

Kintsch, W. & Dijk, T.A.van (1978). Toward a model of text comprehension and reproduction. *Psychological Review*, 85, 363 -394.

Lebiere C., Wallach D., & Taatgen N. (1998). Implicit and explicit learning in ACT-R. In F.E. Ritter & R. M. Young (Eds.), *Proceedings of the Second European Conference on Cognitive Modelling (ECCM-98)*, pp. 183-189. Nottingham: Nottingham University Press.

Marescaux P.-J., Luc F., & Karnas G. (1989). Modes d´apprentissage selectif et nonselectif et conaissances acquises au controle d´un processus: Evaluation d´un modele simule. *Cahiers de Psychologie Cognitive*, 9, 239-264.

Nosofsky, R.M., Clark, S.E. & Shin, H.J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 282 -304.

Ostergaard, A. L. (1992). A method for judging measures of stochastic dependence: Further comments on the current controversy. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 18, 413-420.
Schoppek W. (1998). Modeling causal induction. Paper presented at the Fifth Annual ACT-R Workshop 1998, Carnegie Mellon University, Pittsburgh, PA. (http://act.psy.cmu.edu/ACT/ftp/workshop/Workshop-98/Schoppek/quick_index.html)
Snodgrass, J. G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. Journal of Experimental Psychology: General, 117, 34-50.
Taatgen, N. (1999). Learning without limits. Groningen, The Netherlands: Universal Press of the Rijksuniversiteit Groningen.
Tulving, E., & Hayman, C.G. (1993). Stochastic independence in the recognition/identification paradigm. European Journal of Cognitive Psychology, 5, 353-373.
Vollmeyer R., Burns B.D., & Holyoak K.J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. Cognitive Science, 20, 75-100.