Title
Intervention in Multiple-Cue Judgment: Not Always for the Better

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN
1069-7977

Authors
Enkvist, Tommy
Juslin, Peter
Newell, Ben R.
et al.

Publication Date
2005

Peer reviewed

Intervention in Multiple-Cue Judgment: Not Always for the Better

Tommy Enkvist (tommy.enkvist@psyk.uu.se)
Department of Psychology, Uppsala University, SE-751 42, Uppsala, Sweden

Peter Juslin (peter.juslin@psyk.uu.se)
Department of Psychology, Uppsala University, SE-751 42, Uppsala, Sweden

Henrik Olsson (henrik.olsson@psyk.uu.se)
Department of Psychology, Uppsala University, SE-751 42, Uppsala, Sweden

Ben R. Newell (ben.newell@unsw.edu.au)
School of Psychology, University of New South Wales, Sydney, Australia

Abstract

Previous studies suggest improved learning when participants actively intervene rather than passively observe the stimuli in a judgment task. In two experiments the authors investigate if this improvement generalizes to multiple cue judgment tasks where judgments may be formed from abstract knowledge of cue-criterion relations or exemplar memory. More specific hypotheses were that intervention in learning should improve performance over observation, and that improvement should be associated with a relative shift from exemplar memory to cue abstraction. In contrast to previous studies, in a multiple-cue judgment task with binary cues and continuous criterion, there was poorer learning with intervention than observation, and participants actively experimenting more produced poorer judgments. The results suggest that intervention may distract from efficient exemplar encoding and improvement may be limited to tasks efficiently addressed by cue-abstraction.

Introduction

Information about our environment is acquired in a variety of ways. We learn by instruction from others, by passive observation, and by acting on our environment, observing the consequences of our interventions. These methods of acquiring information are all fundamental to our ability to adapt and function successfully in the environment. In this paper we examine the effects of different learning activities in a multiple cue judgment task and, specifically, how these activities relate to the cognitive representations that support the judgment process. Will people's knowledge of a judgment task differ depending on how they interact with stimuli in training? The learning method used in numerous multiple cue judgment and categorization tasks is observation, where stimuli are presented; the participants make a judgment, and receive outcome feedback. This method is relatively passive and participants cannot interact with the stimuli and actively test their own hypotheses. There is relatively little work examining the role that intervention might have for learning and how these effects relate to cognitive representations.

In the Western scientific tradition the advantage of experimentation over simple observation has long been recognized (Mill, 2002), but in the areas of categorization and multiple-cue judgment, research on intervention is scarce. Existent research on intervention in learning suggests that, perhaps not surprisingly, learning is promoted by the possibility to causally intervene with the system under study (Klayman, 1988; Lagnado & Sloman, 2004). Advances in the understanding and modeling of causal relations (Pearl, 2000) has stimulated renewed interest in cognitive science for casual reasoning and its role for learning (Gopnik et al., 2004; Rehder, 2003; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). Lagnado and Sloman (2004), for example, used a trial by trial based learning paradigm in which participants obtained probabilistic data about a causal chain either through observing sequences (e.g., *seeing* a high fuel temperature and a low combustion chamber pressure leading to the launch of a rocket) or through intervention (e.g., *setting* both temperature and pressure to either high or low and then observing whether a rocket launched or failed). The results showed a clear advantage for interveners in terms of their ability to subsequently select the causal model likely to have generated the data (from an array of possible models).

In this article we extend this finding by investigating the role of intervention in multiple-cue judgment. In a multiple-cue judgment task participants use a number of cues to infer a criterion. Research with this task (Juslin, Jones, Olsson, & Winman, 2003; Juslin, Olsson, & Olsson, 2003) has identified two qualitatively different cognitive processes that can underlie performance in the standard observation training regime. The first, inspired and motivated by research on categorization, emphasizes exemplar memory and assumes that people make judgments by retrieving similar exemplars from memory (Medin & Schaffer, 1978; Nosofsky & Johansen, 2000). The second, derived from research on multiple-cue judgment, stresses the controlled integration of explicit knowledge of cue-criterion relations abstracted in training (Einhorn, Kleinmuntz, & Kleinmuntz, 1979).

In this context we investigate the relationship between active intervention and the cognitive representations acquired and we will argue that the benefit of intervention should be especially large if one engages in cue abstraction. Accordingly, one hypothesis in regard to why intervention affords a benefit over observation is that it promotes knowledge representation in the form of abstract cue-criterion relations rather than memory for exemplars (Juslin, Jones, et al, 2003;

Juslin, Olsson, et al., 2003). Because causal intervention allows controlled observation, for example, by keeping all cues but one constant to investigate its effect on the criterion, arguably it should become easier to abstract the relations between individual cues and the criterion. One prerequisite for this benefit from intervention is that people spontaneously engage in this sort of "experimentation".

But is intervention beneficial in all tasks, also tasks where judgment is supported by processes other than analysis of causal or functional relations between cues and criterion? Previous studies have mainly relied on tasks that invite use of analytic thinking and abstraction. In this experiment we use a simple additive task with many repetitions of the same small set of stimuli. This task can be solved either by abstraction of cue-criterion relations or exemplar memory, and apparently is (Juslin, Olsson, et al., 2003). Note that while intervention is beneficial for inferring the task structure and abstracting the cue-criterion relations, in regard to exemplar memory it may incur a cost. A consequence of cue abstraction is that participants may concentrate more on the piecemeal analysis of individual cue-criterion relations at the cost of considering and encoding entire feature patterns. Therefore, encouraging cue abstraction may have the side-effect of promoting less efficient encoding of the exemplars.

With this background in mind, the experiments in this paper investigated whether the benefit from intervention generalizes also to a multiple cue judgment task where both cue abstraction and exemplar memory are viable processes. We also ascertained whether the participants spontaneously engage in the sort of controlled observation or "experimentation" afforded by intervention. Finally, to the extent that we observed a benefit from intervention, we wanted to test the hypothesis that this benefit is mediated by a relative shift from exemplar memory to reliance of cue abstraction.

Models and Judgment Task

The two processes that are perhaps most often discussed in categorization learning are rule-based and exemplar-based processes (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Juslin, Jones et al., 2003; Juslin, Olsson et al., 2003; Sloman, 1996; Smith, Patalano, & Jonides, 1998). Ruled-based models, like the *cue abstraction model* (CAM), implements the idea that people use controlled processes in working memory to mentally integrate cues according to a linear additive rule. In training participants abstract cue weights that are used to estimate the criterion when a new probe is presented (Juslin et al., 2003). In contrast, *exemplar models* (EBM) assume that people make judgments by retrieving similar stored exemplars from memory (Medin & Schaffer, 1978; Nosofsky & Johansen, 2000), a process that involves rapid similarity-based processes. The exemplars retrieved from memory are representations of holistic concrete experienced instances encountered in training.

We rely on an experimental design that has been successful in distinguish between cue abstraction and exemplar memory in a multiple cue judgment task (Juslin, Olsson et al., 2003). The task involves a probe defined by four binary cues and requires a judgment of a continuous criterion. Judgments are initially made in a training phase where feedback about the correct criterion is provided after every judgment. The cover story involves judgments of the toxicity of subspecies of the exotic (but fictitious) Death Bug.

The task is summarized in Table 1. The cues C_1, C_2, C_3 and C_4 take on values 1 or 0, where the value 1 signifies an increase in toxicity. The toxicity c of a subspecies is a linear additive function of the cues:

$$c=50+4\cdot C_1+3\cdot C_2+2\cdot C_3+1\cdot C_4 \qquad (1)$$

The criterion c is thus computed by assigning cue number one, C_1, most importance and therefore the largest weight and cue number four, C_4, the least importance.

When participants make judgments of the continuous criterion CAM suggests that they perform a mental analogue of linear multiple regression. For each cue, the weight ω_i (i=1…4) is retrieved and the estimate of c is adjusted accordingly:

$$\hat{c}_R = k + \sum_{i=1}^{4} \omega_i \cdot C_i \,, \qquad (2)$$

where $k = 50 + .5 \cdot (10 - \sum \omega_i)$. The value of k is to ensure a rotation of the intercept about 55. If ω_1=4, ω_2=3, ω_3=2, and ω_4=1, Equations 1 and 2 are identical and, in principle, the CAM affords perfectly accurate judgment in this task.

Table 1. All 16 exemplars and their binary cue values. *T* and *O* are exemplars viewed both under training and test. *N* and *E* are new exemplars only presented in the test phase.

Exemplar		*Cues*			Criteria	
#	C_1	C_2	C_3	C_4	Add	Role
1	1	1	1	1	60	*E*
2	1	1	1	0	59	*T*
3	1	1	0	1	58	*T*
4	1	1	0	0	57	*O*
5	1	0	1	1	57	*N*
6	1	0	1	0	56	*N*
7	1	0	0	1	55	*N*
8	1	0	0	0	54	*T*
9	0	1	1	1	56	*O*
10	0	1	1	0	55	*O*
11	0	1	0	1	54	*T*
12	0	1	0	0	53	*T*
13	0	0	1	1	53	*T*
14	0	0	1	0	52	*T*
15	0	0	0	1	51	*T*
16	0	0	0	0	50	*E*

EBM implies that the participants make judgments by retrieving similar exemplars from memory. When EBM is applied to judgments of a continuous criterion, the estimate $\hat{c}_E$ of the criterion c is a weighted average of the criteria c_j

stored for the J exemplars, where the similarities $S(p,x_j)$ are the weights:

$$\hat{c}_E = \frac{\sum_{j=1}^{J} S(p,x_j)\cdot c_j}{\sum_{j=1}^{J} S(p,x_j)}. \qquad (3)$$

where p is the probe to be judged, x_j is stored exemplar j (j=1…J), $S(p,x_j)$ is the similarity between probe p and exemplar x_j. Eq. 4 is the original *context model* (Medin & Schaffer, 1978) applied to a continuum (see, Delosh et al., 1997; Juslin et al., 2003). The similarity between probe p and exemplar x_j is computed according to the multiplicative similarity rule of the original context model:

$$S(p,x_j) = \prod_{i=1}^{4} d_i, \qquad (4)$$

where d_i is an index that takes value 1 if the cue values on cue dimension i coincide (i.e., both are 0 or both are 1), and s_i if they deviate (i.e., one is 0, the other is 1). s_i are four parameters in the interval [0, 1] that capture the impact of deviating cues values (features) on the perceived similarity $S(p,x_j)$. A value of s_i close to 1 implies that a deviating feature on this dimension has no impact on the perceived similarity and is considered irrelevant. A value of s_i close to 0 means that the similarity $S(p,x_j)$ is close to 0 if this feature is deviating, thus assigning crucial importance to it. For low s_i, only identical exemplars have an effect on the judgments.

Extra- and interpolation measures how well the correct cue-criterion relations have been abstracted (DeLosh, 1997). If the participant is able to make correct judgments for the new extreme exemplars presented in the test phase (i.e. is able to extrapolate), the participant is likely to have figured out the underlying cue-criterion relations. This suggests the use of cue abstraction. If an exemplar process is used the participants are unable to extrapolate beyond the range of stimuli seen in training. When judging the new exemplars in the middle range, there will be no systematic differences between new and old with a cue abstraction process. With exemplar processes old exemplars are more correctly judged than new exemplars, because old exemplars with the correct criterion can be retrieved from memory.

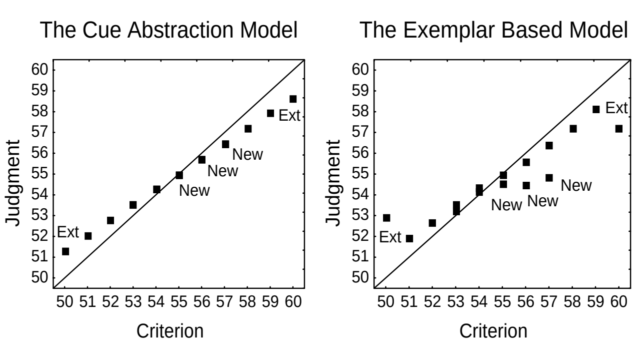

Figure 1. The judgment task design allows for distinct predictions by cue abstraction and exemplar-based processes. *Left:* No systematic inter- or extrapolation effects are expected with cue abstraction process. *Right:* With an exemplar-based process, the participant cannot extrapolate for the new extreme exemplars and the middle range judgments are better for old than new for new exemplars.

Experiment 1

The aims of Experiment 1 were three-fold: First, to replicate the benefit of active intervention observed in previous studies with the sort of multiple-cue judgment used in the present study. Second: to investigate if people spontaneously realize this possibility to make controlled observation when they can actively intervene with the stimuli in training? Third: provided that the beneficial effect of intervention is replicated, to find out if this improvement is associated with a shift from exemplar memory to cue abstraction.

In Experiment 1 we compared the performance by participants who observed (the *observation condition*) or actively constructed (the *intervention condition*) the stimuli in the training phase. Although interveners are given a target criterion for the exemplar they create by selecting the cue values, they do have the freedom to make controlled observations so as to make cue abstraction more efficient, for example, by structuring the successive exemplars that they create so that they differ with respect to a single cue. This should promote cue abstraction, while observation with no control over the cues should make cue abstraction harder; in particular, if it is heavily biased towards estimating linear slopes between successive exemplars differing with respect to a single cue (Juslin, Karlsson, & Olsson, 2004).

We hypothesized that interveners should exploit the possibility to make controlled observations by more often creating successive exemplars differing with respect to only one cue as compared to the baseline provided by the observers, and that, relative to the observers, there should be a shift from exemplar memory to cue abstraction among interveners. The possibility to make controlled observation should improve learning for interveners and we expected them to need fewer trials to reach a certain learning criterion.

Method

Participants. Forty-eight undergraduate students from Uppsala University volunteered. All received payment of approximately 80 SKr ($10) or course credit. Thirty-eight participants were women and ten were men. The participants had a mean age of 25.5 years (range 19-44, SD=5.66). All participants were tested individually.

Materials and Procedure. For the observers each learning trial consisted of the presentation of text descriptions of a fictitious death bug species with four binary attributes (long or short legs, green or brown back, long or short nose, spotted or unspotted fore back). Five exemplars were omitted from training. The omitted bugs were Exemplars 1, 5, 6, 7 and 16 (see Table 1). All participants were shown the exemplars in a new and independent random order.

The participants answered the question “What is the toxicity of this bug?” and made a numerical response (% toxicity) on each trial. Feedback about the correct toxicity followed the response and remained on the screen until the participant clicked to advance to the next trial. A minimum of 45 trials was completed and then training continued until the participant had satisfied a pre-determined learning criterion (an average Root Mean Square Deviation RMSD of 0.8 or less between judged and actual toxicity over the preceding 11 trials).

Interveners saw the same screen layout as observers but rather than seeing predetermined configurations of attributes, interveners selected four attributes on each trial (e.g., brown back, long nose, short legs, spotted fore back). Their task was to create a bug of a given toxicity. Thus, on a trial they were asked, for example, to "Create a bug that has toxicity 57%". After selecting all four features and clicking "create" feedback on the toxicity of the created bug was given. The program did not allow participants to create the five omitted exemplars, ensuring that in both conditions only the Old exemplars in Table 1 could be seen. If an omitted exemplar were about to be created an error message appeared on the screen saying "An error occurred in creation, please try another configuration".

Consistent with the observation condition participants completed a minimum of 45 trials and then continued until achieving the same predetermined criterion. Following learning half of participants from both groups received a test phase in which each of the 16 exemplars was presented twice in a random order, for a total of 32 trials. Participants responded in the same manner as observation learning, but received no feedback. The other half from both groups received an intervention test phase, where the task was to create a bug of a given toxicity without any feedback. This partition was made to see if same training and test condition could boost performance (Morris, Bransford, & Franks, 1977). Finally, all participants were given a series of questions and tasks designed to elicit their insight into the structure of the task (not reported in this article).

Results and Discussion

The results from the test phase are analyzed in two different ways. First, in terms of performance where judgment accuracy in the two conditions is compared; second, by cognitive modeling, examining how CAM and the EBM are manifested in the participant's judgment through a model fit.

Performance. A two-way ANOVA on RMSE with training condition (observation vs. intervention) and test condition (observation vs. intervention) as between-subjects factors produces no statistically significant effect of test ($F_{1, 44}$= .39, p= .54), no significant interaction ($F_{1, 44}$= .98, p= .33), but a marginally significant effect of training ($F_{1, 46}$= 3.32, p= .08). Observation in training produced more accurate judgments than intervention (a RMSE of 1.21 vs. 1.56). The number of training trials needed to reach the training criterion was similar ($F_{1, 46}$= .14, p= .72).

An Experimentation Index (EI) is calculated for the 40 first trials in the training phase. The EI in the intervention condition tells weather participants has experimented, how many cues that are held constant from trial to trial, during training. More cues held constant suggests more experimentation. The difference in EI is significant ($F_{1,46}$=20.79, p= .000), implying that interveners do more experimentation than the baseline obtained in observation. The correlation between the EI and RMSE was r_{48} = .19, not significant (p= .2).

Table 2: Mean RMSE, number of training trials needed to reach the training criterion, and EI for both conditions.

Condition	RMSE	Training Trials	EI
Observation	1.21	90.75 (SD=41.8)	1.86
Intervention	1.56	85.92 (SD=49.2)	2.12

Model Fit. The four best-fitting parameters for each model were ascertained by minimizing the SSD between model predictions and the last judgment made for each of the 11 exemplars in the training phase (see Table 1). The models with these parameters were then used to predict how the participants should perform in the test phase with all 16 exemplars (see Juslin et al., 2003). This implies cross-validation for the 11 old exemplars and genuine predictions for the 5 new exemplars. The models thereby predict what the judgments would be like if the participants in each task used either the EBM or the CAM. The fit between the 16 predictions and the 16 mean judgments made in the test phase is measured by the RMSD. Model fit was performed on individual level. A Split-plot ANOVA with training and test condition (observation vs. intervention) as between-subjects factors and model (CAM vs. EBM) as within-subjects factor yielded a significant main effects of training condition ($F_{1,46}$=7.2, p=.001), significant effect of model ($F_{1,46}$=16.57, p=.000), but as illustrated in Figure 2 (left panel) it is the significant interaction ($F_{1,44}$=11.74, p=.001) that is the key effect. In the observation training condition the CAM provides superior fit, but in the intervention training condition both models show relatively poor fit.

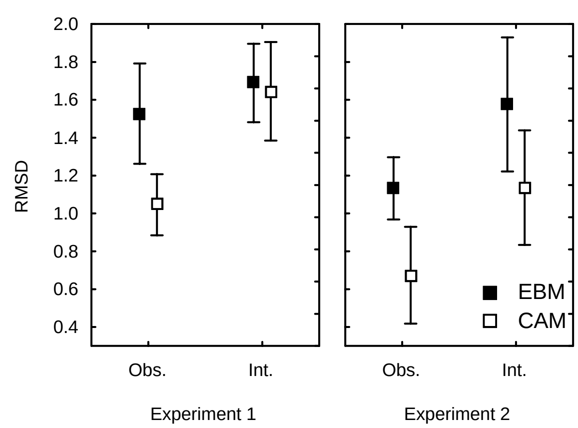

Figure 2. Mean RMSD with 95% confidence intervals for the cue abstraction and the exemplar models in the observation and intervention conditions.

Experiment 1 revealed no benefit in learning for participants that learned the task by active intervention. Instead there was a strong tendency towards more accurate judgments and better fit for CAM with observation. Both groups needed a similar number of trials to reach the pre-determined learning criterion. It seems clear that these data provide no evidence for increased performance with active intervention or for a shift towards more cue abstraction, despite that the participants in the intervention condition spontaneously engaged in "experimentation" when this was made possible.

Experiment 2

To further investigate if active intervention with stimuli in training produces a representational shift between different representations Experiment 2 was conducted. The training phase in earlier experiments with this paradigm has been 220 trials long (Juslin, Jones et al., 2003; Juslin, Olsson et al., 2003). One possibility is that in the intervention condition, the participants begin with more or less fragmentary exemplars and as training proceeds cue abstraction is slowly improved by more extensive experience with intervention.

The aim of Experiment 2 was therefore to investigate if increasing the number of training trials would increase the difference between the two conditions in regard to performance and dominating representation. In addition, Experiment 2 serves the complementary purpose of collecting more data on one intriguing aspect of Experiment 1; the marginally significant deterioration in performance for interveners. Manipulation of observation versus intervention at test produced no effects in Experiment 1 and in Experiment 2 observation was used in the test phase for both conditions.

Method

Experiment 2 followed the same procedure as Experiment 1. The difference between the two experiments was the number of learning trials. Rather than continuing until reaching a predetermined learning criterion, all participants completed 220 trials, ensuring that each of the 11 exemplars was presented (Observation) or asked for (Interveners) 20 times. Twenty-four undergraduate students from Uppsala University took part and were rewarded in the same manner as in Experiment 1. Ten participants were male and 14 were female, with a mean age of 25.71 (range 20-45, SD=4.85).

Results and Discussion

The data in Experiment 2 were analyzed in the same way as in Experiment 1. Comparisons are made between Experiments 1 and 2 to investigate the effect of increased training.

Performance. A one-way ANOVA on the RMSE shows a marginally significant difference, again favoring passive observation over active intervention ($F_{1,\ 22}$= 3.22, p=.09). The experimentation indices was computed for separate blocks of trials in the training phase, this time for 11 blocks of 20 successive trials were entered as the dependent variable in an ANOVA. The ANOVA shows a significant main effect of condition ($F_{10,\ 242}$ = 34.5, p = .000), a non-significant main effect of block ($F_{10,\ 242}$ = .64, p = .78), and a non-significant interaction ($F_{10,\ 242}$ = .67, p = .75), see Figure 3. The correlation between the Experimentation index and RMSE was significantly positive (r_{24} =.5, p = .014), suggesting that more experimentation was associated with *poorer* judgment accuracy.

Table 3. Mean RMSE and mean EI in Experiment 2.

Condition	RMSE	EI
Observation	.76	1.9
Intervention	1.24	2.1

Model Fit. Model fit was analyzed in the same way as in Experiment 1. The model fit in Experiment 2 in terms of mean RMSD when the models were fitted to the individual participant data is summarized in Figure 2 (right panel). A Split-plot ANOVA with training condition (observation vs. intervention) as between-subjects factor and model (cue abstraction vs. exemplar memory) as Within-subjects factor show a significant main effect of model ($F_{1,22}$ = 17.26, p = .000) and of training ($F_{1,22}$ = 10.16, p = .004), but no significant interaction ($F_{1,22}$ = .01, p = .93). As illustrated in Figure 2, both models show better fit for the observation than the intervention condition, and the cue abstraction model shows superior fit in both of these training conditions.

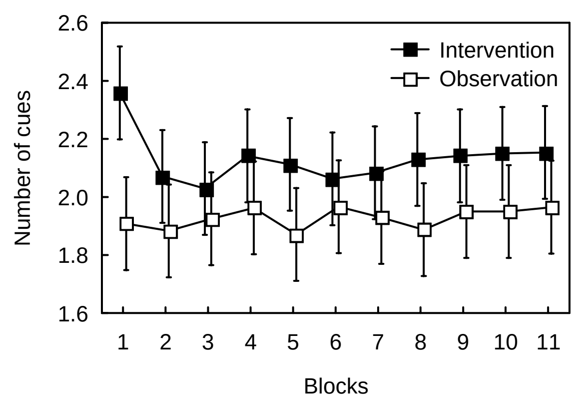

Figure 3. The mean Experiment Index with 95% confidence intervals for both conditions in Experiment 2 plotted as a function of training block, where each block consists of 20 trails. The means for observation define the change of cues expected by chance.

Experiment 2 provides further evidence that active intervention can actually instill poorer learning than passive observation in a task with binary cues, where more active experimentation actually contributes to poorer judgments. We also performed an analysis with data aggregated across Experiments 1 and 2. A two-way ANOVA with RMSE as dependent variable and intervention vs. observation and Experiment 1 vs. 2 as independent variables shows significant main effects of training condition ($F_{1,\ 68}$=6.33, MSE=.44, p=.01) and Experiment ($F_{1,\ 68}$=5.6, MSE=.44, p=.02), but no statistically significant interaction ($F_{1,\ 68}$=.15, MSE=.44, p=.7). The more extensive training in Experiment 2 produced more accurate judgments and intervention contributed to poorer judgments as compared to passive observation. The increased training did not serve to separate the two conditions from each other and there were no signs of a shift from exemplar memory to cue abstraction. The overall superiority of cue abstraction was even clearer in this experiment as compared to the previous experiments and this superiority does not appear strengthened by allowing the participants to actively intervene with stimuli.

General Discussion

In two experiments we examined if different activities in the learning phase affect the knowledge representation in a multiple cue judgment task. Previous studies suggests that intervention in learning improves performance (Gopnik et al., 2004; Klayman, 1988; Lagnado & Sloman, 2004; Steyvers et al., 2003). In our experiments we have found that intervention and causal experimentation does not always improve performance. Apparently, in an environment with binary cues and a continuous criterion intervention as a learning method can not compete with simple and passive observation. A possible explanation is that improvement in judgment with intervention as learning method is limited to environments that spontaneously invite cue abstraction. When cues are binary and fewer exemplars exist in the environment, memorization of exemplars precedes the abstraction of rules in both learning conditions. The use of intervention and the search of a rule suppresses the possibility to memorize correct exemplars, but in observation the process of memorization of exemplars starts immediately.

No shift in representation from EBM to CAM could be found. In the observation condition CAM had significantly better fit in both experiments. In intervention no differences existed between CAM and EBM in Experiment 1, but in Experiment 2 CAM became the best fitting model.

The poor performance in the Intervention condition (especially in Experiment 1) raises intriguing issues about how exemplars are coded and stored in memory. A default assumption in research on perceptual classification is that exemplars are stored in terms of visual features of the objects, as is naturally the case in the Observation condition. In the Intervention condition it is however possible that, exemplars are not exclusively or even mainly coded in terms of visual features, but in terms of the sequence of actions required to produce an exemplar of a specific toxicity. Clearly, the exact interpretation of “exemplar” is open to somewhat different interpretations in this case.

To conclude, our results suggest that the proposition that experimentation is better than simple observation is not always true. Contrary to Mill and the believers in the superiority of active learning, intervention as a learning strategy is not always better.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442-481.

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 968-986.

Einhorn, J. H., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Regression models and process tracing analysis. *Psychological Review, 86*, 465-485.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 1-30.

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning Memory and Cognition, 29*, 924-941.

Juslin, P., Karlsson, L., & Olsson, H. (2004). *Additive integration of information in multiple cue judgment: A division of labor hypothesis*. Manuscript submitted for publication.

Juslin, P., Olsson, H., & Olsson, A. C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General, 132*, 133-156.

Klayman, J. (1988). Cue discovery in probabilistic environments: Uncertainty and experimentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 317-330.

Lagnado, D. A., & Sloman, S. A. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning Memory and Cognition, 30*, 856-876.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Mill, J. S. (2002). *A System of Logic: Rationcinative and Inductive*. (first edition published in 1843): University Press of the Pacific.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin and Review, 7*, 375-402.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University Press.

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1141-1159.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65*, 167-196.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27*, 453-489.