Title
Non-linear Multiple-Cue Judgment Tasks

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

ISSN
1069-7977

Authors
Enkvist, Tommy
Juslin, Peter
Olsson, Anna-Carin

Publication Date
2005

Peer reviewed

Non-linear Multiple-Cue Judgment Tasks

Anna-Carin Olsson (anna-carin.olsson@psy.umu.se)
Department of Psychology, Umeå University
SE-091 87, Umeå, Sweden

Tommy Enqvist (tommy.enqvist@psyk.uu.se)
Department of Psychology, Uppsala University
Box 1225, 751 42 Uppsala, Sweden

Peter Juslin (peter.juslin@psyk.uu.se)
Department of Psychology, Uppsala University
Box 1225, 751 42 Uppsala, Sweden

Abstract

In two experiments we examined which cognitive processes people use in a non-linear multiple cue judgment task, proposing that people are not able to use explicit cue abstraction when judging objects with a non-linear structure between the cues of the objects and the criterion and therefore they are forced to use exemplar-based processes. Consistent with the results reveal strong exemplar effects in the non-linear condition.

Introduction

How do we judge if there seems to be no underlying structure or pattern in the things or events we have to judge? What knowledge representations do we use in these situations? Our ability to categorize the environment is important to us and different environmental categorization situations involve different processes, either in the form of analytic or intuitive thought. Multiple cue judgment tasks typically involve a probe defined by a number of binary or continuous cues and require a continuous judgment. Previous research suggests that by manipulating the task environment we can induce a shift between distinct cognitive processes (Juslin, Olsson, & Olsson, 2003). Two cognitive models of general concern in cognitive science, are the explicit, rule-based and analytic knowledge versus implicit, silent knowledge based on personal experience (Hahn & Chater, 1998; Hammond, 1996; Sloman, 1996; E. E. Smith, Patalano, & Jonides, 1998). Research on multiple cue judgment has primarily stressed controlled integration of cues that have been abstracted in training (Einhorn, Kleinmuntz, & Kleinmuntz, 1979) while exemplar models are emphasized by research on categorization (Nosofsky & Johansen, 2000; Nosofsky & Palmeri, 1997). In this article we address the following question: what structure of a categorization task triggers analytic thinking and what structure triggers intuition? We hypothesize that participants are not able to use cue abstraction in a non-linear task, because the mind is constrained to linear additive integration of abstracted cue-criterion relations (Juslin, Karlsson, & Olsson, 2004).

Cognitive models and processes

In general, multiple cue judgments are well captured by multiple linear regression models (Brehmer, 1994, Cooksey, 1996). These regression models are statistical descriptions rather than process models, but implicitly this research is often committed to the idea that people use controlled processes in working memory to mentally integrate cues according to a linear additive rule. In contrast, exemplar models assume that people make judgments by retrieving similar stored exemplars from memory (Medin & Schaffer, 1978; Nosofsky & Johansen, 2000). The exemplar representations in memory are the holistic concrete experienced instances often emphasized in categorization.

The generalized judgment model Σ implicates a sequential adjustment process, compatible both with a linear, additive cue-integration rule, and the additive combination of exemplars in exemplar models (Juslin et al., 2004), capturing the additive character of a judgment. By considering the structural properties of a task environment, we are able to predict representational shifts in the process that support multiple cue judgment. In additive environments cue abstraction dominate, while exemplar memory dominates in multiplicative environments (Juslin et al., 2004).

In the Σ model the previous estimate is adjusted every time a new piece of evidence is presented. If no new evidence is presented cues correspond to an a priori estimate. In Σ, cue abstraction involves sequential adjustment based on single cues, where the cue weight is relative to the cues presented so far. In contrast, the weight assigned to an exemplar is given by its similarity relative to the similarity to all exemplars retrieved so far. Because of the integration of stored criterion values instead of cue-criterion relations, Σ is not confined to any particular task structure. Therefore, the model can represent any task structure as long as similar exemplars have similar criteria, which allows judgments also in non-linear and multiplicative environments (Juslin et al., 2004).

Judgment task and Model Predictions

The participants used four binary cues to infer a continuous criterion in either a linear or a non-linear multiple-cue judgment task. The exotic (but fictitious) Death bug is the cover story in the task that involves judgments of the toxicity of subspecies of the bug, which can be inferred from four cues of the subspecies (leg length, nose length, spots or no spots on the fore back and different patterns on the buttock). The concentration of poison varies from 50 to 60 ppm in each subspecies (Jones et al., 2000; Juslin et al., 2003; Juslin et al., 2004). The task structure is summarized in Table 1.

Table 1: Structure of the judgment task with the constrained training set for the linear and non-linear conditions.

Exemplar	Cues				Criteria		Role
#	C1	C2	C3	C4	Linear	Non-linear	Constr. Tr. set
1	1	1	1	1	60	50	E
2	1	1	1	0	59	53,6	T
3	1	1	0	1	58	56,4	T
4	1	1	0	0	57	58,4	O
5	1	0	1	1	57	58,4	N
6	1	0	1	0	56	59,6	N
7	1	0	0	1	55	60	N
8	1	0	0	0	54	59,6	T
9	0	1	1	1	56	59,6	O
10	0	1	1	0	55	60	O
11	0	1	0	1	54	59,6	T
12	0	1	0	0	53	58,4	T
13	0	0	1	1	53	58,4	T
14	0	0	1	0	52	56,4	T
15	0	0	0	1	51	53,6	T
16	0	0	0	0	50	50	E

Note: T= training exemplar; O= training exemplar that serves as old exemplar in the interpolation comparison; E= new exemplar that only occurs in the test phase for measuring extrapolation; N= new exemplar that only occur in the test phase for interpolation comparison.

The values of the binary cues take on 1 or 0. The function of the cue values to judge the toxicity c_L of a subspecies is linear and additive in the linear judgment task:

$$c_L = 50 + 4 \cdot C_1 + 3 \cdot C_2 + 2 \cdot C_3 + 1 \cdot C_4 \tag{1}$$

A quadratic function of the criteria in the linear condition was made as the function in the non-linear judgment task with the same range (50 to 60 ppm) and maximum value (60 ppm) with a non-linear relationship between the cues and the criterion instead of the linear function.

$$c_{NL} = -2 \cdot c_L{}^2/5 + 44 \cdot c_L - 1150 \tag{2}$$

The functions c is the level of poison in the bug and $C_{1 \ldots} C_4$ are weighted binary features of the bug. For example a bug with features [1101] has a toxicity level of 58 in the linear condition and a toxicity level of 56.4 in the non-linear condition, see Table 1. The criterion c is computed by assigning the most important cue, C_1, and has the largest weight with the least important cue that has the smallest weight, C_4. If the binary cue has value 1, it suggests high toxicity level, and if the cue has value 0, it suggests low toxicity level, see Table 1. In Experiment 1 a normally and independently random error was added. The variance of the random error would produce a .9 correlation (see Juslin et al, 2003). Figure 1 illustrates the relationship between the two functions in a plotted diagram. The non-linear function is a more complex environment than the linear function. The increase of complexity in a function or environment is when a function goes from an additive linear function to be a multiplicative non-linear one. When a participant can not use a simple additive rule to calculate what correct criterion is in a judgment. The non-linear function can also be plotted against the criterion values for the non-linear function on the x-axis (Figure 2), where the function only has six criterion values and two or four examples matching each criterion (see Table 1).

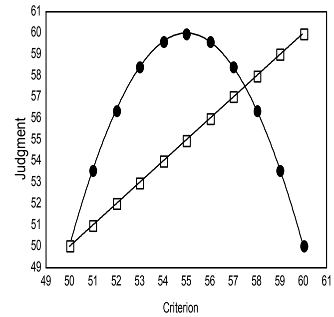

Figure 1. The additive linear function and the non-linear function plotted against the criteria in the linear condition. Both functions have the same range on the x-axis (criterion) and the same range on the y-axis (judgment).

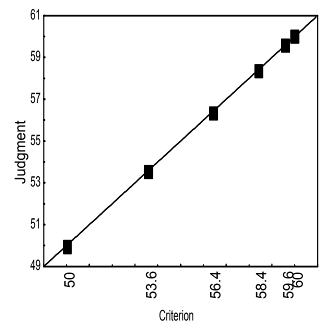

Figure 2. The non-linear function plotted against the criterion values for the non-linear function on the x-axis, with only six criterion values, where two or four examples matching each criteria.

Cognitive Models

Cue abstraction model. The application of the cue abstraction model to continuous criteria implies that participants abstract cue weights ω_i, which specify the importance and the sign of the relation of the cue (i=1…4). In training participants abstract the cue weights and use those to compute an estimate of the criterion when a new probe is presented. The estimate of c is adjusted according to the cue weight and the final estimate $\hat{c}_R$ of c is a linear additive function of the cue values C_i. For example, after training the rule for cue C_1 may specify that C_1 goes with a large increase in toxicity. This corresponds to the standard application of a linear additive equation to model multiple cue judgment,

$$\hat{c}_R = k + \sum_{i=1}^{4} \omega_i \cdot C_i \tag{3}$$

where $k = 50 + .5 \cdot (10 - \sum \omega_i)$. If ω_1=4, ω_2=3, ω_3=2, and ω_4=1, Equations 1 and 2 are identical and the model produces perfectly accurate judgments. The intercept k constrains the function relating judgments to criteria to be regressive around the midpoint (55) of the interval [50, 60] specified by the task instructions.

Exemplar model. The application of exemplar model to a continuous criteria we assume that judgments are made by retrieving similar exemplars from memory and the estimate 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,

$$\hat{c}_E = \frac{\sum_{j=1}^{J} S(p,x_j)\cdot c_j}{\sum_{j=1}^{J} S(p,x_j)} . \qquad (4)$$

Eq. 5 is the context model (Medin & Schaffer, 1978) applied to a continuum (see Delosh et al., 1997; Juslin & Persson, 2000; E. R. Smith & Zarate, 1992). The application of an exemplar model to multiple-cue judgment is illustrated in Figure 3. The similarity between probe p and exemplar x_j is computed according to the multiplicative similarity rule of the original context model (Medin & Schaffer, 1978):

$$S(p,x_j) = \prod_{i=1}^{4} d_i , \qquad (5)$$

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 overall perceived similarity $S(p,x_j)$. s_i close to1 implies that a deviating feature on this cue dimension has no impact on the perceived similarity and is considered irrelevant. 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 the feature. The parameters s_i capture the similarity relations between stimuli and the attention paid to each cue dimension, where a low s_i signifies high attention. In effect, for low s_i, only identical exemplars have a profound effect on the judgments. For example, with all s_i=.001 identical exemplars receive weight 1 in Eq. 4, but exemplars with just one deviating feature receive weight .001. With s_i close to 1, on the other hand, all exemplars receive the same weight, regardless of the number of deviating features.

Predictions by the Models

Predictions by the cue abstraction and exemplar model are summarized in Figure 3. When participants are trained with all 16 exemplars it is impossible to discriminate the different models from each other, because they predict the same accurate judgments (see Figure 3, panel A and B). This perfect accuracy depends on correct knowledge of the cue weights and error-free integration of this knowledge into a judgment in the cue abstraction model while it derives from retrieval of stored exemplars, where only identical exemplars are allowed to have a strong effect on the judgment in the exemplar model.

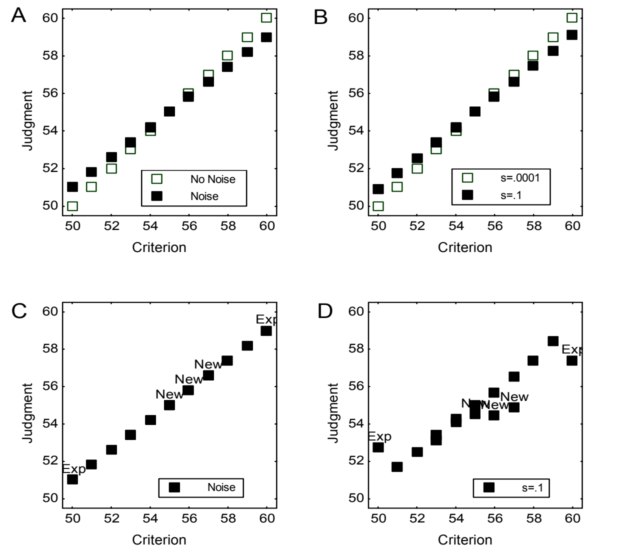

Figure 3: Predictions for the continuous task. Panel A: Cue abstraction models with no noise and noise for the complete training set. Panel B: Exemplar model with all similarity parameter s equal to .0001 and .1 for the complete set. Panel C: Cue abstraction model with noise for the constrained set. Panel D: Exemplar model with similarity parameter s=.1 for the constrained set.

The discrimination of the models involves the existence or no existence of *extrapolation* and *interpolation*, in other words, the ability to make accurate judgments of new exemplars. When a constrained set of subspecies are omitted in training phase (two extreme exemplars, e.g. [1 1 1 1] and [0 0 0 0] and three middle exemplars, [1 0 1 1], [10 1 0], [1 0 0 1], see Table 1 for new N, and old, O exemplars) and the participants are presented with the complete set of subspecies in the test phase, the cue abstraction model affords extrapolation and interpolation, because the model allows the adding function of integrated cue values that produce accurate judgments. The participants will figure out that the extreme subspecies with all cues presented and all cues absent should have the most extreme value, even if they have never been presented to these subspecies in the training phase, also no old-new differences between the intermediate exemplars omitted in training should exist in the cue abstraction model. In contrast, the exemplar model is not able to extrapolate or interpolate, when the responses for the new exemplars are determined by retrieval of identical exemplars (Delosh et al., 1997; Erickson & Kruschke, 1998). Because the exemplar model involves linear combination of the criteria observed in training that range between the toxicity levels 51 and 59, it can never produce a judgment outside this range, as extrapolation requires. The exemplar model also predicts old-new differences with more accurate judgments for old exemplars than for new when the toxicity level can be retrieved from memory (Juslin et al., 2003).

Method

Participants

Fifty-four undergraduate students from the university participated in the two experiments (thirty participants in Experiment 1 and twenty-four participants in Experiment 2). Twenty-three males and twenty-seven females. The average age was 24.25 years. Each participant received 70 SEK (approximately 8$) in payment for participation in the study.

Design and procedure

The design of the two experiments was a between-subjects design, where all participants made the judgments in training and test with same stimuli presentation format. Each experiment contained two conditions each. In Experiment 1, a probabilistic linear judgment task condition and a probabilistic non-linear judgment task condition

The written instructions informed the participants that there were different subspecies of a Death Bug and that the task was to estimate the toxicity (poison level) of the subspecies as a number between 50 and 60. The experiments contained two phases, where the first phase was a training phase which provided trial-by-trial outcome feedback about the continuous criterion ("This bug has toxicity 56.7%"). In the training phase 11 training exemplars of the bug were presented 20 times each, a total of 220 trials with criterion values between 51 and 59 in the linear condition and between 53.6 and 60 in the non-linear condition. The remaining five exemplars were omitted in the training phase and first presented in the test phase. Participants were trained and tested with analogue stimuli in the form of pictures of the bug species, presented on the computer screen.

The subspecies varied with regard in four binary cues; leg length (short or long), nose length (short or long), spots or no spots on the fore back and two different patterns on the buttock and different colors were used for the cue values to strengthen their salience. The abstract cues in Table 1 were randomly assigned to new visual features for each participant. The cue values had the weights 4, 3, 2 and 1 which determine the portion of toxicity that each cue adds to the total amount. The question asked on the computer screen was "What is the toxicity of this subspecies".

In the test phase, all 16 exemplars were presented, including the five exemplars omitted in the training phase. The test phase went over 32 trials, where the 16 exemplars were presented twice in a random order. The participants made same judgments as in the training phase but received no outcome feedback. The whole experiment took 45 minutes.

Experiment 2 involved the same stimuli and non-linear structure as Experiment 1. However, in contrast to Experiment 1, the cue criterion-relation was changed from a probabilistic task to a deterministic task. The length of the training phase was also changed from 220 trials to 440 trials. The instructions were the same in the first condition (Old-instructions condition) as in Experiment 1 but was changed in the second condition (New instructions condition), and encourage the participants to memorize each bug exemplar. We expected the determinism and the exemplar-memorization request to increase the performance and the use of exemplar memory.

Dependent Measures

The dependent measures reported involve: performance, the representation index, and model fit. The performance measures are, Root Mean Square Error (RMSE) of judgments (between judgment and criteria), and consistency (correlation between the two judgments made for the same exemplar in the test phase).

A Representation index (RI) was calculated to see what level of representation that dominates the judgments in the two different tasks. The interpolation measure is obtained by taking the difference between absolute deviation between judgment and criteria for old exemplars and new exemplars as computed by the following formula,

$$I = \sum \left[d(Old)_n - d(New)_n \right] / 6 \qquad (6)$$

Where n is refers to criterion 55, 56 and 57 in the linear condition and criterion 58.4, 59.6 and 60 in the non-linear condition denoted either as "Training(Old)" or "Interp.(New)" in Table 1. The exemplars in the extrapolation are also considered. Extrapolation is measured by the deviation from an expected linear extrapolation, based on mean judgment of the old exemplars in the interpolation range. The difference between actual judgment and expected value is the extrapolation measure. A 0 in the extrapolation index implies that the judgements for the extreme exemplars are as extreme as the expected regression-based extrapolation for the old exemplars, and the judgments are correct by all linear transformations of the correct judgments. This suggests appropriate extrapolation. When extreme exemplars do not receive as extreme judgments as expected from extrapolation and the index is negative. This in turn, supports the exemplar model and its inability to extrapolate. In other words, the exemplar model does not allow accurate extrapolation while cue abstraction does (see Figure1, for a concrete illustration).

To increase the statistical power in the analysis of the data and for ease of exposition of the data the interpolation index and extrapolation index are combined which gives us the RI. The RI indicates if participants have based their judgments on exemplar model or cue abstraction. A RI of 0 implies the regression-based paradox with the ability to extrapolate while a negative index implies the inability to extrapolate, exactly the same as mentioned before, in the extrapolation index.

By using the data from the training phase in each task to create predictions of the two models was performed in a computer simulation. The Model fit measures are the coefficient of determination (r^2) and Root Mean Square Deviation (RMSD) between predictions and the computed test phase data.

Results Experiment 1

Performance. The probabilistic linear task was lower (0.56) than for the probabilistic non-linear task (0.85)in consistency, with a significant difference ($F(1.28)=4.61$; $p=0.04$; $MSE=0.137$). The errors in the judgment as measured by RMSE (Root Mean Square Error) was significant different in performance of the two tasks ($F(1.28)=37.0$; $p=0.0000$; $MSE=0.44$), where the probabilistic additive linear task showed a lower RMSE value (1.54) than the probabilistic non linear task (3.02). In Figure 4 the test mean judgment from the probabilistic additive linear task and the probabilistic non-linear task are plotted against the correct criteria.

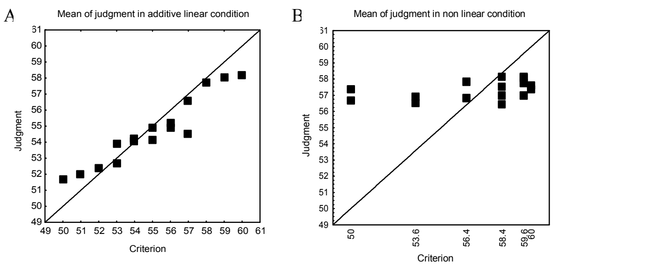

Figure 4. Test mean judgment and exemplar index for the probabilistic additive linear task (Panel A) and the probabilistic non-linear task (Panel B) plotted against the correct criteria. The full drawn lines show the correct judgment.

Representation. The linear judgment task has a lower RI value (-0.67) than the non-linear judgment task (-1.33), but not significant different ($F(1, 28)=0.79$; p=0.38; $MSE=4.08$). The RI in the non-linear condition is significantly separated from zero but not in the linear condition. This suggests that EBM is the model in use in the non-linear condition, while both CAM and EBM is in use in the linear condition.

Model fit. The models in Eq. 3 and 4 were fitted to the mean judgments computed for the constrained training set with 11 subspecies of the bug across the last 110 trials. Cue abstraction model allows analytic derivation of the best-fitting parameters that corresponds to logistic regression. By the Quasi-Newton method in the MathCAD software the parameters for the exemplar model with the minimal squared sum of error were obtained.

The model fits (see Figure 5) show that in the probabilistic additive linear condition both EBM and CAM have high correlation indices and about the same RMSD values (EBM; r^2= .92 and RMSD=0.55 and CAM; r^2= .96 and RMSD= 0.54). In the non linear condition EBM and CAM presents approximately the same RMSD (EBM; RMSD= 0.51 and CAM; RMSD = 0.59) as in the linear condition, while the correlation is very low (EBM; r^2= .49 and CAM; r^2= .24), with some advantage for EBM.

When looking at the training data we see that the curve of learning was very flat, which imply that the participants may not have learned to asymptote in the training phase. This could has effected the results and therefore Experiment 2 was constructed to provide a better setting for testing the hypothesis that participants are not able to use cue abstraction in a non-linear task.

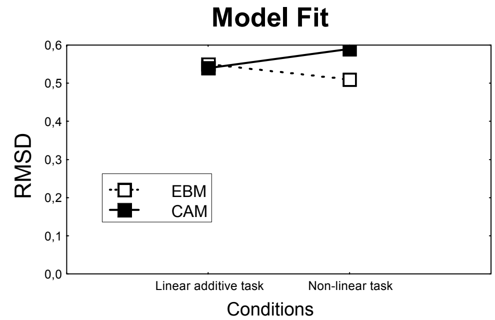

Figure 5. Model fits of exemplar model and cue abstraction model in the linear additive task and the non-linear task.

Results Experiment 2

Performance. The consistency for the two conditions was nearly the same (.17 for the regular instruction condition and .18 for the exemplar instruction condition). The RMSE showed a lower value for the exemplar instruction condition (1.93) than the regular instruction condition (2.50), with significant difference (F(1.10)=9.28; p=0.002; MSE=6.70). In Figure 6 the mean test judgment from the regular instruction condition and the exemplar instruction condition are plotted against the criteria.

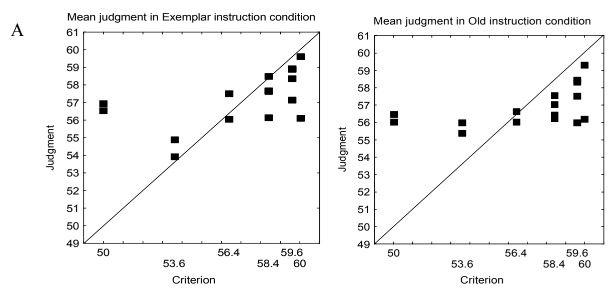

Figure 4. Test mean judgment and exemplar index for the regular instruction condition (Panel A) and the exemplar instruction condition (Panel B) plotted against the correct criteria. The full drawn lines show the correct judgment.

*Representation.*The RI shows that the regular instruction condition has a higher RI value (-2.68) than the exemplar instruction condition (-2.79), but the difference is not significant (F(1.10)=0.30; p=0.86; MSE=2.40). The RI is separated from zero which suggests that EBM is used in both conditions.

Model fit. The same procedure for model fit was used as in Experiment 1. The model fits (see Figure 6) show that EBM has high correlation in both regular instruction condition (EBM; r2=.68 and RMSD=1.03) and exemplar instruction condition (EBM; r2=.87 and RMSD=0.7) compared to CAM in the conditions (regular instruction condition, CAM; r2=.46 and RMSD=1.09 and exemplar instruction condition; CAM; r2=.39 and RMSD=1.27) which suggests a dominated use of EBM in both condition, but strongest in the exemplar instruction condition.

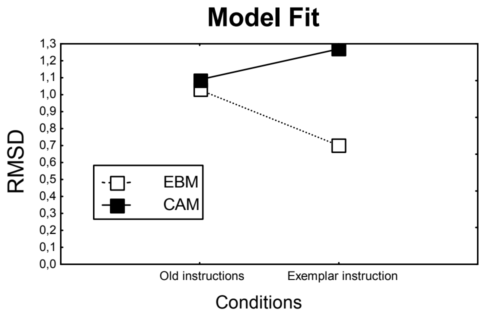

Figure 7. Model fits of exemplar model and cue abstraction model in the old instruction condition and the exemplar instruction condition.

Discussion

The question addressed in this article is what structure of a categorization task triggers analytic thinking and what structure triggers intuition? We hypothesized that participants are not able to use the cue abstraction model, because it is too difficult to induce the equation that underlies the criterion. Overall, the results in Experiment 1 the non-linear condition shows little fit to the rule-based model CAM. The higher RI, the poor results on old and new exemplars in the interpolation and extrapolation range, all support EBM. The results of the RMSD that is almost identical to the linear condition can not suggest which model that is used, if any of the models was. One possible explanation of the low RMSD and the poor consistency correlation is that the judgment curve has a flack inclination, and the task is thus to complex to learn. Because of the lack of learning for the participants in Experiment 1, we investigated in Experiment 2 if longer training phase and a change from a probabilistic to a deterministic non-linear task could allow the participants to use any of the processes. The results clearly showed the use of EBM in both conditions suggesting that determinism and more training make it possible to adopt exemplar-based process and use it more effectively. The use of cue abstraction seems to be non-existent in the experiments which support the hypotheses that the non-linear judgment task is too complex for judgments based on mental integrated cues and the idea that exemplar-based processes will serve as a back-up system when cue abstraction is impossible.

Acknowledgements

Bank of Sweden Tercentenary Foundation supported this research.

References

Brehmer, B. (1994). The psychology of linear judgment models. *Acta Psychologica*, *87,* 137-154.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. Academic Press: San Diego.

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.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127, 107-140.

Hahn, U., & Chater, N. (1998). Similarity and rules: distinct? Exhaustive? Empirically distinguishable? *Cognition, 65*, 197-230.

Jones, S., Juslin, P., Olsson, H., & Winman, A. (2000). Algorithm, heuristic or exemplar: Process and representation in multiple-cue judgment. In L. Gleitman, & A. K. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 244-249). Hillsdale, NJ: Erlbaum.

Juslin, P, & Karlsson, L., Olsson, H. (2004) Exemplar memory in Multiple-cue Judgment: A Division of Labor Hypothesis (in press)

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

Juslin, P., & Persson, M. (2003). PROBabilities from Exemplars (PROBEX): A lazy algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, *26*, 563-607.

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

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

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266-300.

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