# Learning in Multiple-cue Judgment Tasks

**Bettina von Helversen (Bettina.vonhelversen@unibas.ch)**
University of Basel, Department of Psychology, Missionsstr, 62a
4057, Basel, Switzerland

**Jörg Rieskamp (joerg.rieskamp@unibas.ch)**
University of Basel, Department of Psychology, Missionsstr, 62a
4057, Basel, Switzerland

**Abstract**

In our daily lives we often make quantitative judgments based on multiple pieces of information such as evaluating a student's paper based on form and content. Psychological research suggests that humans rely on several strategies to make multiple-cue judgments. The strategy that is used depends on the structure of the task. In contrast, recent research on learning in judgment tasks suggests that learning is relatively independent of task structure. In a simulation study we investigated how the performance of several learning models is influenced by the structure of the task and the amount of learning experience. We found that a linear additive neuronal network model performed well regardless of task structure and amount of learning. However, with little learning a heuristic model performed similarly well, and with extensive learning, associative learning models caught up with the linear additive model.

**Keywords:** Learning; multiple-cue judgments; Computational modeling

## Multiple-cue Judgments

When judging objects on a continuous criterion such as the quality of a research paper, people often rely on multiple sources of information. For example, the clarity of the writing, the novelty of the research and the methodological precision may be used as important aspects for evaluating a paper. Several models have been developed to describe how humans solve these judgment problems. Traditionally, linear additive models have been employed to capture how humans weigh and integrate information. Social Judgment Theory (SJT; see Doherty and Kurz, 1996; Cooksey, 1996) relied on multiple-linear regression models to capture decision policies and researchers have used this approach successfully to describe judgments in many areas (see Brehmer, 1988). Similarly, Anderson (1981) suggested that humans combine information in a linear additive fashion. However, recently it has been suggested that humans may have multiple cognitive strategies available to make multiple-cue judgments. Juslin, Karlsson, and Olsson (2008) suggested that depending on the structure of the tasks, humans may switch between a rule-based cue abstraction approach and a similarity–based exemplar approach. Similarly, von Helversen and Rieskamp (2008, 2009) suggested the mapping model, a heuristic model for multiple-cue judgments, and showed that the model that was best in describing participants' behavior depended on the task structure. More specifically, they showed that the mapping model described participants' responses well in tasks that could not be solved by a linear model and where participants had knowledge about the cues' polarity; that is, the sign of the correlation between a cue and the criterion. The exemplar model performed well, in non-linear environments with no prior knowledge about cue polarity, and a linear additive model performed well if the task structure was linear.

## Learning in Multiple-cue Judgment Tasks

Although many studies in multiple-cue judgment research rely on extensive learning phases, there have been relatively few attempts to understand and model the learning process. However, the learning process is crucial to understand how people come to make judgments and which cognitive processes they rely on. Particularly, if — as suggested — people rely in their judgment on multiple cognitive processes, this should also be reflected in the learning phase. Additionally, the learning phase itself could play an important role in determining how later judgments are made. Recently, Kelley and Busemeyer (2008) compared how well several models could describe the learning process in various multiple-cue judgment tasks. They compared a rule-based neuronal network model with a delta-learning rule (e.g. Gluck & Bower, 1988), which can be seen as a learning version of a linear additive model with an associative connectionist network model (ALM, Busemeyer, Byun, DeLosh, & McDaniel, 1997; Busemeyer, Myung, & McDaniel, 1993). They found that the rule-based neuronal network models described the learning process best in the majority of the tasks, suggesting that learning may be relatively independent of task structure.

These results are somewhat contrary to the research by Juslin et al. (2008) and von Helversen and Rieskamp (2009) on multiple-cue judgments, suggesting that humans rely on a variety of strategies, depending on the structure of the task (e.g. Juslin, et al., 2008; Rieskamp & Otto, 2006). This raises the question of whether learning depends on the task structure and what may be the mechanisms that lead to a switch in cognitive processing during learning. In this paper we investigate two reasons that may cause a shift in cognitive processing during learning in a multiple-cue judgment task. One reason to rely on different learning strategies may be that their learning performance differs depending on the structure of the task. Thus, we will

investigate if task structure influences how well various learning procedures perform that are imbedded in different cognitive models of multiple-cue judgments (e.g. Juslin et al., 2008; Kelley & Busemeyer, 2008; von Helversen & Rieskamp, 2008). Second, the reliance on different learning procedures could also be due to differences in how fast the procedures adapt to different judgment structures. Therefore, we additionally examined if the models differ with respect to their learning speed.

## Learning Models

We tested learning versions of cognitive models suggested in the literature for multiple-cue judgments. As a learning model for the linear additive model we relied on a rule-based neuronal network model as implemented by Kelley and Busemeyer (2008). As an exemplar model we extended the ALCOVE model (Kruschke, 1992) to continuous judgments. ALCOVE has been successfully used to model exemplar-based learning in categorization. We also tested a version of the mapping model (von Helversen & Rieskamp, 2008) to allow for learning. Additionally, we included the ALM model as implemented by Kelley and Busemeyer (2008).

**Linear Additive Model** Much research has shown that linear additive models are good at describing human judgments (Brehmer, 1994). The linear additive model assumes that people weigh each piece of information according to its importance and then add the weighed evidence to reach a judgment. Traditionally, a multiple linear regression is used to capture how much weight people put on each piece of information (i.e. cue). Kelley and Busemeyer (2008) used a rule-based neuronal network with a linear additive structure:

$$g_t = a_1 \cdot c_1 + a_2 \cdot c_2 + ... + a_k \cdot c_k, \qquad (1)$$

where the model prediction $g$ at time $t$ is given by the sum of the cue values $c$ for $k$ cues weighted by their importance $a$ at time $t$. This learning model updates the weight for each cue according to a delta rule (Gluck & Bower, 1988) with a learning parameter δ capturing the learning rate. An additional decay parameter ω controls the impact of new information.

$$a_{k,t} = a_{k,t-1} + (\delta / t^{\omega})(Y_{t-1} - g_{t-1})c_{k,t-1}, \qquad (2)$$

with $Y$ indicating the feedback (i.e. the criterion value) and $g$ the model prediction at time $t$-1.

**Mapping model** We extended the mapping model (von Helversen & Rieskamp, 2008) to allow for learning. The mapping model follows a simple cognitive strategy that makes judgments by first categorizing an object and then retrieving a typical estimate for the category it was put in. According to the mapping model, an object is placed into a category based on the sum of (standardized) cue values, implying that all cues are weighted equally. The judgment is then determined by the median of the criterion values of all objects in the respective cue sum category. The learning procedure we suggest describes how and how many cue sum categories are formed during learning. In the beginning it is assumed that only a single category is used. In each learning trial, the decision is then made as to whether the new object is put into a new category or into an existing category. A new category is formed if the difference between the cue sum of a new object and the cue sum of each existing category is larger than a distance parameter $d$. The criterion value estimated for each category is the mean of the criterion values of the objects falling into this category and is updated whenever a new object falls within a category.

**ALM** The ALM model is an associative connectionist network model. It assumes a layer of input nodes representing each combination of cue values (2^Number of cues, with binary cue data). The input nodes are connected to a layer of $r$ output nodes reflecting the criterion values via a one-dimensional grid of equally spaced values. Input nodes are activated by a stimulus based on the similarity of the stimulis' cue values $C$ to the input node's cue values $I$.

$$A_t = \exp(-\gamma(C_t - I_t)^2), \qquad (3)$$

with the activation $A$ of the input nodes at time $t$ further depending on a scaling parameter γ that determines the slope of the activation gradient. The activation of the input nodes is spread to the output nodes via connection weights. The activation of an output node $O_r$ is given by the sum of activations of the input nodes weighted by the connection weights between the input nodes and the output node. The probability of choosing an output node is given by the ratio of the activation of the output node to the summed activation of all output nodes. The judgment is a weighted average of the output nodes, where each output node is weighted with the probability with which it is chosen. Connection weights are updated at each trial according to a delta-learning rule. For this it is assumed that the feedback criterion value produces a feedback activation of each output node $F_r$ based on the similarity of the feedback value $p_t$ to the output node $p_r$:

$$F_r(p_t) = \exp(-\gamma(p_t - p_r)^2). \qquad (4)$$

The connection weights α are updated based on the feedback activation $F$, the predicted activation $O$ and the input activation $A$, with a learning parameter δ capturing the learning rate:

$$\alpha_t = \alpha_{t-1} + \delta\left[F_{t-1} - O_{t-1}\right]A_{t-1}. \qquad (5)$$

**ALCOVE** We extended ALCOVE (Kruschke, 1992) to continuous judgments. ALCOVE has a similar structure as

the ALM model; however, the input nodes of ALCOVE are restricted to the exemplars encountered during learning. As in ALM the activation of an input node is based on the similarity of the stimulus object to the input node. However in ALCOVE, similarity depends also on the attention given to each cue dimension $k$, which is controlled by a set of attention weights $w$.

$$A = \exp(-\gamma \sum_k w_k [c_k - i_k]^2), \quad (6)$$

with the activation $A$ of an input node based on the squared distance of the stimulus value $c$ on dimension $k$ to the value of the input node $i$ on cue dimension $k$, weighted by the attention $w$ given to this cue dimension and a scaling parameter γ determining the slope of the activation gradient. In the original ALCOVE model, one output node is chosen as response. To allow for continuous judgments we extended ALCOVE with the ALM's estimation mechanism described above.

In ALCOVE, the connection weights are updated in the same way as in ALM, with learning parameter $\delta_1$ capturing the learning rate (see Equations 4 and 5). Additionally, the attention weights are also updated according to a delta learning rule. The learning rate is captured in an additional free parameter $\delta_2$. The attention weights $w$ are updated according to the following rule:

$$w_{k,t} = w_{k,t-1} - \delta_2 \gamma \sum_r \left[ (F_r - O_r) \sum_n A_n \alpha_{n,r} (c_k - i_k)^2 \right], \quad (7)$$

with $r$ indexing the output nodes, $n$ the input nodes and $k$ the cue dimensions; $F$ gives the respective feedback activation and $O$ the predicted activation of an output node. $A$ indicates the respective activation of an input node, $\alpha$ is the connection weights between the input and output node and $c_k$ and $i_k$ provide the stimulus value and the input node value on cue dimension $k$.

## Method

To test how the performance of the learning models in solving judgment tasks depend on the task structure, we compared the models' performance by computer simulations in two environments: a linear environment and a multiplicative environment. Furthermore, we varied the amount of learning to examine the relationship between the models' performance and the size of the training set.

**Simulation Environments** We created two different environments: a linear environment and a multiplicative environment similar to the environments used by von Helversen and Rieskamp (2008; Experiment 3), which revealed a strong effect of task structure on people's judgment processes. Each environment consisted of 1000 objects described by 5 binary cues, with randomly drawn values (0 or 1). The criterion in the linear environment $Y_L$ was generated by a linear additive function:

$$Y_L = 30 + 33c_1 + 22c_2 + 20c_3 + 15\,c_4 + 5c_5 + \varepsilon. \quad (8)$$

The error term ε was drawn from a normal distribution with a mean of zero and a standard deviation of 10. The multiplicative criterion $Y_M$ was generated by a multiplicative function:

$$Y_M = 1.2 \cdot \exp(Y_L / 30), \quad (9)$$

resulting in criterion values with similar ranges (about 0 to 140) in both environments.

**Simulation Procedure** For the simulation we drew a random training sample of 250 objects 50 times and a hold-out set of 100 from each of the environments. Then we fitted the free parameters of the four models to the training data minimizing the square deviation between the model prediction and the training data. For the linear additive model we assumed that in the beginning, equal weight would be given to all cues. For the associative models we assumed that the connections weights and attention weights had equal starting values. Based on the estimated parameter values we generated model predictions for the hold-out set after seeing 20, 50, 150 and 250 objects from the training set. As a measure of prediction accuracy we calculated the root mean square deviation (*RMSD*) between the model prediction and the criterion in the hold-out set after the four points of learning and averaged across the trials of the simulation at each point of learning. Since parameters are fit on a separate data set, the performance of the models in the hold-out set can be compared without needing to further adjust for the complexity of the models.

## Results

The mean best fitting parameter values for the models are reported in Table 1, indicating similar learning in both environments.

Table 1: Mean parameter values (SD)

| | Environment | |
|---|---|---|
| Parameters | Linear | Multiplicative |
| Linear additive: δ | .45 (.30) | .30 (.17) |
| Linear additive: $\omega$ | .45(.14) | .47 (.13) |
| Mapping: $d$ | 0 (0) | .02 (.14) |
| ALCOVE: γ | .30 (.36) | .22 (.17) |
| ALCOVE: $\delta_1$ | .42 (.56) | .46 (1.44) |
| ALCOVE: $\delta_2$ | 145 (50) | 173 (63) |
| ALM: γ | 2.72 (.31) | 1.78 (.30) |
| ALM: δ | .14 (.07) | .22 (.07) |

The models differed with regard to how well they learned the criterion values in the training set. In particular, the two

associative models performed less well than the mapping model and the linear additive model (see Table 2).

Table 2: Mean model performance in *RMSD* (SE) in the training set

| | Environment | |
|---|---|---|
| Models | Linear | Multiplicative |
| Linear additive | 11.09 (.07) | 9.78 (.21) |
| Mapping | 14.60 (.08) | 9.87 (.16) |
| ALCOVE | 15.18 (.09) | 10.32 (.18) |
| ALM | 15.05 (.12) | 11.51 (.17) |

The results in the hold-out set suggest that the performance differences in the training set are partly due to a slow initial learning process of the associative models. Figures 1 (linear environment) and 2 (multiplicative environment) show that the linear additive model and the mapping model learn rather quickly even with as little as 20 learning trials. However, the two associative models that performed worse with less than 50 learning trials caught up with the other two models after extensive learning of 150 trials or more.

The environment crucially influenced the performance of the models. Unsurprisingly, in the linear environment, the linear additive model performed best regardless of the amount of training. With fewer than 50 learning trials, the mapping model performed somewhat worse than the linear model, but better than the associative models. However, with more than 150 trials of learning the two associative models performed better than the mapping model and almost as good as the linear additive model.

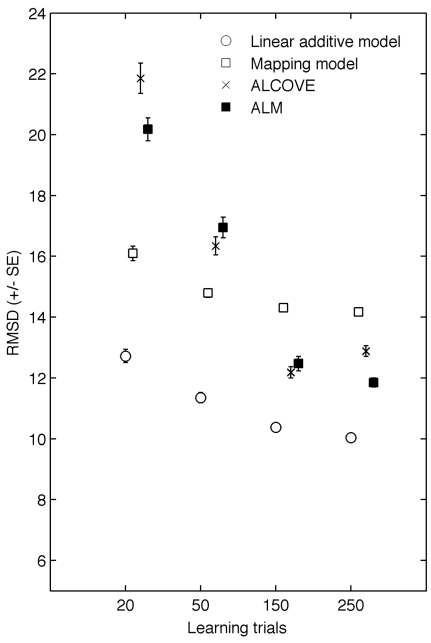


Figure 1: Model performance (*RMSD*) in the hold-out set in the linear environment after 20, 50, 150 and 250 trials of learning. Error bars denote one standard error of the mean.

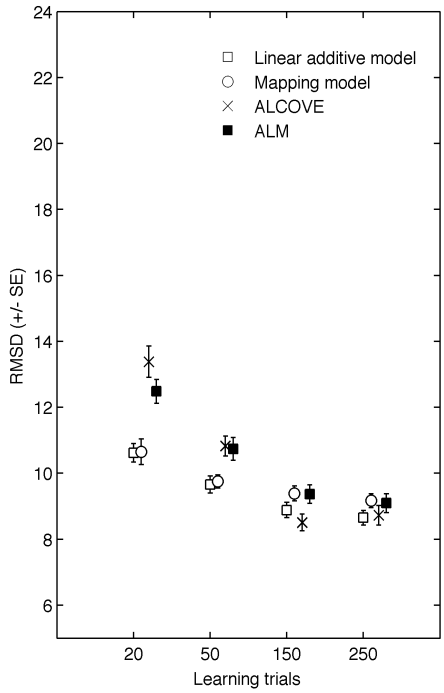


Figure 2: Model performance (*RMSD*) in the hold-out set in the multiplicative environment after 20, 50, 150 and 250 trials of learning. Error bars denote one standard error of the mean.

In the multiplicative environment, the advantage of the linear additive model was less pronounced. To begin with, it performed equally well as the mapping model, but gained a bit on the mapping model with more than 150 trials of learning. The two associative models again performed worse than the linear and the mapping models with little learning with fewer than 50 learning trials, but caught up after more than 150 trials of learning.

In summary, the linear additive model performed well in both environments and at all stages of learning. Furthermore, we found evidence that the amount of training affected which models are well suited to making accurate judgments. More specifically, the associative models only made accurate judgments after extensive training. In contrast, the mapping model performed reasonably well with little training, but failed to improve to a similar degree as the other models with further training.

## Discussion

We investigated how different learning models can solve a multiple-cue judgment task depending on the amount of learning and the structure of the task. We found that a linear additive neural network model performed well in both environments and regardless of the amount of training. However, we also found differences due to task structure. In the multiplicative environment, the mapping model was

equally as good as the linear additive model, in particular with little learning experience. With extensive learning experience the two associative models, ALCOVE and ALM, performed similarly well to the linear additive model and the mapping model. The results are in line with the finding of Kelley and Busemeyer (2008) that a neural network with a linear basis was well suited to describe participants' judgments over a broad range of tasks. Our results also support research illustrating the robust performance of linear models for judgment tasks (Hogarth & Kareleia, 2007).

However, our results seem to contradict results that suggest task-dependent changes in strategy use in multiple-cue judgments (Juslin, et al., 2008; von Helversen & Rieskamp, 2008, 2009). These authors found in a task with a similar structure as in our simulation, that the model that was best in describing participants' judgments changed depending on the task structure. However, the judgment process people rely on might not only depend on the judgment performance of the learning process (e.g. Ashby, Alfonso-Reese, Waldron & Turken, 1998). Instead, the learning speed and also other factors such as the cognitive effort of relying on a specific cognitive process could also influence which judgment and learning process people follow (see also Enkvist, Newell, Juslin, & Olsson, 2006). Particularly, in the multiplicative environment the mapping model may provide an equally good but arguably cognitively simpler alternative, which could explain why a majority of participants were best described by the mapping model in the multiplicative condition of Experiment 3 by von Helversen and Rieskamp (2008). On the other hand, associative processes seem to provide a valid alternative to a linear additive model after extensive training, in particular in a multiplicative environment. If following the assumption that associative similarity-based processes may be executed without conscious awareness and be thus cognitively less demanding (e.g. Ashby & Maddox, 1994), this could still make it attractive for participants to rely on such processes, particularly after extensive training. This could explain the reliance on exemplar-based processes (Juslin, et al., 2008) and also the considerable minority of participants that were best described by the ALM model (see Kelley & Busemeyer, 2008).

Lastly, the available context information may also influence people's strategy choices. Information about cue polarity seems to trigger rule-based processes (Newell, Weston, Tunney, & Shanks, 2009; von Helversen & Rieskamp, 2009). While in the study by Juslin et al., (2008) participants had no information about cue polarity, most studies analyzed by Kelley and Busemeyer (2008) provided context information that allows drawing conclusions about cue polarity and thus could have increased the reliance on rule-based processes.

## Conclusion

In sum, our results suggest that linear additive learning models are generally robust. However, the performance advantage depends on the task structure and the amount of learning opportunity. On the basis of these results future research will test whether people's judgments depend on task characteristics and learning opportunities.

## Acknowledgments

This research was supported by a grant of the German Research Foundation to the first and second author (RI 1226/5).

## References

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press

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.

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

Brehmer, B., & Joyce, C. R. B. (1988). *Human judgment: The SJT view*. Amsterdam: Elsevier/North Holland.

Busemeyer, J.R., Myung, I.J., & McDaniel, M.A. (1993) Cue competition effects: Theoretical implications for adaptive network learning models. *Psychological Science,* 4, 196-202

Busemeyer, J. R., Byun, E., DeLosh, E. L., & McDaniel, M. A. (1997). Learning functional relations based on experience with input-output pairs by humans and artificial neural networks. In K. Lamberts & D. Shanks (Eds.), *Knowledge concepts and categories* (pp. 405–437). Cambridge, MA: MIT Press.

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

Doherty, M. E., & Kurz, E. M. (1996). Social judgment theory. *Thinking and Reasoning, 2*, 109-140.

Enkvist, T., Newell, B. R., Juslin, P., & Olsson, H. (2006). On the role of causal intervention in multiple-cue judgment: Positive and negative effects on learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 163-179.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*, 227-247.

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: matching rules and environments. *Psychological Review, 114*, 733.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition, 106*, 259-298.

Kelley, H., & Busemeyer, J. (2008). A comparison of models for learning how to dynamically integrate multiple cues in order to forecast continuous criteria. *Journal of Mathematical Psychology, 52*, 218-240.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.

Newell, B., Weston, N., Tunney, R., & Shanks, D. (2009). The effectiveness of feedback in multiple-cue probability

learning. *The Quarterly Journal of Experimental Psychology, 62*, 890-908.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General, 135*, 207-236.

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General, 137*, 73-96.

von Helversen, B., & Rieskamp, J. (2009). Models of quantitative estimations: Rule-based and exemplar-based processes compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 867-889.