UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
A Matter of Process Accuracy: Observing or Inferring the Criterion of Few or Many Exemplars

Permalink
https://escholarship.org/uc/item/8rp0g3vs

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 34(34)

ISSN
1069-7977

Author
Henriksson, Maria

Publication Date
2012

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

A Matter of Process Accuracy: Observing or Inferring the Criterion of Few or Many Exemplars

Maria P. Henriksson (maria.henriksson@psyk.uu.se)
Department of Psychology, Uppsala University
SE-751 42, Uppsala, Sweden

Abstract

Can we tailor fit the training to enhance judgment accuracy by changing to the learning format that invites the most effective cognitive process for the task environment at hand? The results from a study on multiple-cue judgments revealed that observing the cues and the criterion of exemplars simultaneously with no feedback involved in the training, a learning format predicted to invite exemplar memory processes, was the better learning option when there were few unique exemplars in training. Inferring the criteria of different exemplars and receiving outcome feedback during training, a learning format predicted to invite cue-abstraction, was the better learning option when there were many unique exemplars in training. Implications for the notion of an initial "rule bias" suggested by several previous studies are discussed.

Keywords: rule-bias; observation; feedback; cue-abstraction; exemplar memory

Introduction

Virtually all research on multiple-cue judgment has involved the learning format *feedback learning* and multiple-cue learning has often, more or less explicitly, been regarded as an analytic or rule-based process, where outcome feedback is used to adjust cue weights and to test hypotheses about the cue-criterion relations (Klayman, 1988). The *Cue-Abstraction Model* (CAM: e.g. Juslin, Karlsson & Olsson, 2008) is a cognitive model capturing many of these properties of the judgment process, assuming explicit knowledge about cue-weights and a controlled integration of information by an additive rule (see the Method section for more information about this model). If the analytic, abstract knowledge assumed with the CAM accurately reflects (or well approximates) the task environment, the judgments become independent of the concrete exemplars encountered during training. Judgment accuracy for old exemplars experienced in training and new exemplars should thus be similar, with ability to extrapolate the judgments beyond the observed range of training exemplars (DeLosh, Busemeyer, McDaniel, 1997).

However, feedback learning is not the only learning format. In the related domain of category learning there is a growing interest in investigating the effects of *observation learning* where no feedback is involved and people learn from observing the cues and the criterion (see, e.g., Ashby, Maddox, & Bohil, 2002). There is also some evidence that exemplar memory processes can better describe the performance with observation learning than with the standard feedback learning format (Estes, 1994). A recent study on multiple-cue judgments revealed evidence that observation learning invites exemplar processes and is able to exploit more complex task environments with resulting superior performance when the cues are multiplicatively related to the criterion in the task environment (Henriksson, Enkvist & Juslin, 2012). This suggests that exemplar processes might be used by observation learners who only have to store the information about exemplars in memory and use the similarity to these stored exemplars when assessing the criterion value of new exemplars in subsequent judgments. In contrast to the predictions for CAM, exemplar processes predicts that judgment accuracy for old exemplars experienced in training is superior to the judgment accuracy for new exemplars, and that judgments cannot extrapolate beyond the training range of exemplars (Delosh et al., 1997; Medin & Schaffer, 1978; Nosofsky, 1986). See Method section for more information about this model.

In categorization, Rouder and Ratcliff (2004) have found evidence that exemplar processes provides a better account of the data when there are few and distinct exemplars and rule-based processes provides a better account of the data when the exemplars are confusable and not distinct from each other, for example when exemplars are probabilistically assigned to a category. It is conceivable that exemplar processes might be more vulnerable to the number of unique exemplars in the task environment. As the number of different exemplars increases with experience, the memorization might become difficult with interference between exemplars. Thus, the accuracy of exemplar processes might be constrained to task environments where there are a limited number of training exemplars. On the other hand, a cue-abstraction process might require experience of many different exemplars varying on the cue-dimensions for testing and fine-tuning hypotheses about the relative cue-weights and the relationship between cues and the criterion. The prediction for the multiple-cue judgment task is therefore that the better relative fit of EBM (i.e., clearer advantage for EBM over CAM), the better the judgment accuracy when there are few exemplars in training. When there are many training exemplars the prediction is that the better relative fit of CAM the better judgment accuracy.

The recurring "rule bias" (Ashby, Alfonso-Reese & Turken., 1998, p. 467), an initial inclination for analytical processes, is perhaps not surprising considering that the feedback format is often applied in studies on categorization and multiple-cue judgment. It is possible that feedback learning per se invites relatively more cue-abstraction (CAM) or at least reinforces that kind of process. However, it is reasonable to appreciate that exemplar processes can act as a back-up process whenever rule-based processes fails to exploit the task environment (Juslin et al., 2008; Karlsson, Juslin & Olsson, 2008). It is possible that the previous reported shifts to exemplar processes (Ashby et al., 1998; Erickson & Kruschke, 1998; Kalish, Lewandowsky & Davies, 2005) may in part be mediated by a spontaneous shift to observation learning. For example, if the task is difficult to learn by testing explicit hypotheses against feedback, the participant could start to randomly guess the missing value and wait for the correct outcome feedback to appear. Then, the participant will have the same information as an observation learner who only has to store the information in memory for subsequent use.

In sum, the predictions are that with few exemplars, observation is predicted to produce higher accuracy than feedback by inviting the EBM. With many exemplars, feedback learning is predicted to produce higher judgment accuracy than observation by inviting the CAM (see Figure 1 for the predictions).

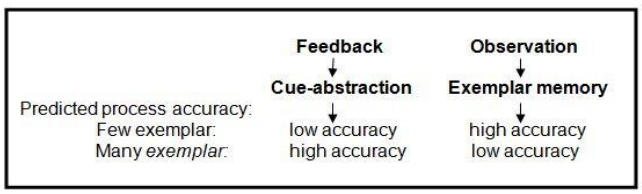

Figure 1. Predicted judgment accuracy and invited processes for feedback and observation learners after training with few or many exemplars.

Method

Participants

Sixty-four students from Uppsala University volunteered. Seven were excluded since their test performance indicated no learning. Of the remaining 57 participants, 40 were women and 17 men with the average age of 24.86 (*SD*= 7.20).

Materials, Procedure, and Design

A computerized multiple-cue learning task was presented to the participants instructing them to learn the suitability for an unspecified job based on values ranging from 0-10 on four cues describing different applicants. The cues were *independent*, *thoughtful*, *detailed-oriented*, and *practical* and they were stated on the computer screen along with the cue-value describing each individual exemplar.

The criterion *c*, the degree of suitability of the exemplar is a linear, additive function of the cues *C*, with the most important cue with a relative weight of .4 and the second-most important cue with a relative weight of .3 and so forth (see Equation 1)[1]. The assignment of the labels of the cues to the relative cue-weights was counterbalanced across the participants.

$$c = 500 + .4 \cdot C_1 + .3 \cdot C_2 + .2 \cdot C_3 + .1 \cdot C_4 \qquad (1)$$

Among the 11^4 possible exemplars that can be generated, two sets of training exemplars were sampled, each with a criterion ranging from 510 to 590. The 16 training exemplars in the condition *few exemplars* were presented 10 times in a randomized order. In the condition *many exemplars* the 16 exemplars were presented only once along with 144 other exemplars in a randomized order (in total 160 training trials in each condition). At test, the 16 exemplars reoccurred together with 14 new exemplars, all with criterion values ranging from 500 to 600. The 30 test exemplars were presented twice in a randomized order. At test, 12 of the old exemplars experienced in training were matched to 12 new exemplars with the same criterion in order to examine new-old differences.

A 2 x 2 factorial design was used and participants were randomly assigned to one of four experimental conditions. The independent variables were the learning format (observation and feedback) and the numbers of exemplars in training (few or many). All participants were told that they were going to learn the degree of suitability of different presented applicants. Half of the participants were told that they should learn by observing the cues and the criterion of different exemplars, similar to screening lists of previous employees' characteristics and degrees' of suitability (observation learning). The other half was told that they should observe the cues describing each individual exemplar and predict the missing criterion value. After each judgment, outcome feedback about the criterion was provided to the participant (feedback learning). Half of the participants in each learning condition experienced few unique training exemplars and the other half experienced many unique exemplars. After the training phase a test phase followed that was identical for all participants. All participants were informed that no feedback should be received during or after this test phase.

The Models and Dependent variables

The *Cue-Abstraction Model (CAM;* e.g. Juslin et al., 2008*)* assumes that the participants abstract cue-weights in training, analogue to linear regression weights. When they later judge the criterion of a probe, they use the knowledge of the cue-weights to integrate the linear additive impact of the cues. For each cue C_i, the weight ω_i (i=1…4) is used when adjusting the criterion $\hat{c}$ of a probe *p*,

[1] Two cues were positively related to the criterion and two cues were negatively related to the criterion so as not to make identification of cue-directions trivial (i.e., with high cue-values always predicting high suitability and low cue-values always predicting low suitability).

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

where the intercept a and the weights ω are parameters in the model.

The *Exemplar-Based Model* (EBM) refers to a version of the generalized context model (Nosofsky, 1986) that is applicable to multiple-cue learning (e.g., Juslin et al., 2008). As many exemplar-based models assume (e.g., Medin & Schaffer, 1978; Nosofsky, 1986), people store memory traces of concrete exemplars together with the outcome. At the time of judgment, people retrieve similar exemplars from long-term memory. According to the Generalized Context Model (GCM: Nosofsky, 1986), the similarity to stored exemplars depends on the attention to the cue dimensions and the sensitivity for the distance between the exemplars in the psychological space. The distance between the probe p and an exemplar j is given by,

$$d_{pj} = h\left[\sum_{m=1}^{4} w_m \left|(x_{pm} - x_{jm})^r\right.\right]^{1/r} \qquad (3)$$

where x_{pm} and x_{jm} are values of the probe and the exemplar on cue dimension m (m=1…4), w_m are attention weights on cue dimension m, and h is a parameter that captures the sensitivity for the distance between the exemplars in the psychological space. The sensitivity varies from 0 to ∞. The attention weights on cues vary between 0 and 1 and are constrained to sum up to 1. Euclidian metric is used and r is set to 2. The overall similarity between a probe p and exemplar j is assumed to be a nonlinear decreasing function of their distance d_{pj} in the psychological space,

$$S(p, x_j) = e^{-d_{pj}} \qquad (4)$$

EBM implies that the criterion $\hat{c}$ of a probe p is assessed by,

$$\hat{c}_p = \frac{\sum_{j=1}^{4} S_j \cdot c_j}{\sum_{j=1}^{4} S_j} \qquad (5)$$

where S_j is the similarity to exemplar j, and c_j is the criterion of exemplar j. The estimated criterion of a probe is the weighted average of the criteria of similar exemplars retrieved from long-term memory, where the similarity is the weight (see Juslin et al., 2008).

This exemplar model and the cue-abstraction model (Juslin et al., 2008) were fitted individually to the responses by each participant in the test phase. A cross-validation procedure was used in the modeling and the model fit is measured by *Root Mean Squared Deviation* (*RMSD*) between the model prediction and the judgment[2]. Judgment accuracy in the test phase is measured by *Root Mean Squared Error* (*RMSE*) between the judgment and the criterion. Hence, the lower value of *RMSE*, the better judgment accuracy. Deltafit (Δfit), a measure of the relative differences in fit of EBM and CAM was computed by subtracting the *RMSD* for CAM from the *RMSD* for EBM so that negative values corresponds to a relatively better fit for EBM and positive values corresponds to a relatively better fit for CAM. Separate analyses of the correlations between the deltafit and judgment accuracy (*RMSE*) can therefore be calculated in order to explore how useful the two cognitive processes are for achieving accuracy when experiencing few or many training exemplars.

Results

A split-plot ANOVA revealed only a significant within-effect of *RMSD* for CAM and EBM, $F(1, 53)=20.82$, $p<.001$. The model with the average best fit was the CAM ($RMSD_{CAM}$=11.10, SD=3.82 vs. $RMSD_{EBM}$=14.2, SD=5.59). The variance explained by the CAM was significantly higher for feedback learners regardless of the number of training exemplars ($r^2_{CAM=}$ .73 and .77; r^2_{EBM}= .61 and .61). Though CAM was found to be the better fitting model for observation by the *RMSD*, the variance explained by the CAM was not significantly higher than for EBM ($r^2_{CAM=}$ .86 and .80; r^2_{EBM}= .77 and .71).

In line with the predictions, a split-plot ANOVA with the judgment accuracy of the matched old and new exemplars at test revealed that there was a significant interaction between the learning formats and the matched exemplars, $F(1,53)=4.76$, p=.034. Observation learners had significantly better accuracy judging the old exemplars that had been experienced in training compared to judging the matched new exemplars. This result suggests that exemplar based processes are used by observation learners. No systematic difference in judgment accuracy between matched old and new exemplars was found for feedback learners, suggesting that cue-abstraction is used by feedback learners (see Figure 2 for illustration of the results).

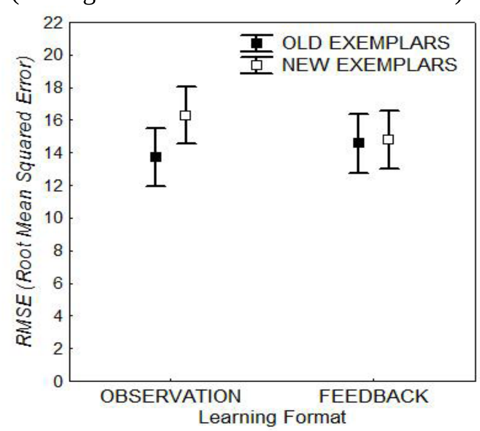

Figure 2. Judgment accuracy (*RMSE*) for the matched old and new exemplars at test for observation and feedback learners after training with few or many exemplars. Lower value of RMSE signifies better judgment accuracy. Vertical bars denote 95% Confidence intervals.

[2] The 2 x 30 judgments of the test exemplars were randomly split into two sets for each participant so that each exemplar occurred once in each set.

In line with the predictions, the deltafit (i.e., the relative fit of the models) had a positive correlation with *RMSE* when there were few training exemplars, r_s =.33, $t(26)$=1.78, n=28, p=.04 one-sided, a result that suggest that the better fit for EBM, the lower the *RMSE* (thus better judgment accuracy). With many experienced exemplars in training, the deltafit had a negative correlation with *RMSE* , r_s = -.34, $t(27)$= - 1.88, n=29, p=.04 one-sided, a result that suggest that the better fit for CAM, the lower the *RMSE* (thus better judgment accuracy). The two correlation coefficients differed significantly (p< .01).

A two-way ANOVA with judgment accuracy (*RMSE*) as dependent variable revealed no main effects of the number of experienced exemplars or learning formats. However, there was a significant interaction effect, F (1, 53) =4.17, p=.046. In line with the prediction, observation learners had marginally significantly better judgment accuracy than feedback learners after training with few exemplars (M=13.41 vs. 15.92, SD=3.69 vs. 5.08, p=.06 by planned comparison). On the other hand, feedback learners had marginally significantly better judgment accuracy than observation learners after training with many unique exemplars (M= 11.58 vs. 13.69, SD= 4.76 vs. 3.39, p=.09 by planned comparisons). As illustrated in Figure 3, the number of exemplars affects more the overall performance for feedback learners than for observation learners. This is consistent with the claim by Juslin et al. (2008) that whenever a cue-abstraction fails to exploit the task environment, it is better to shift to exemplar memory that can act as a back-up process.

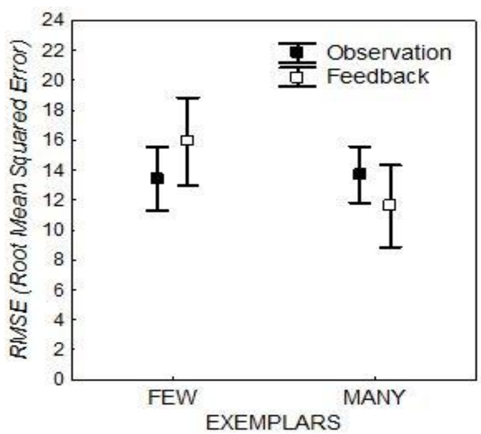

Figure 3. Overall judgment accuracy (*RMSE*) for observation and feedback learners after training with few or many exemplars. Lower value of *RMSE* signifies better judgment accuracy. Vertical bars denote 95% confidence intervals.

Discussion

This study revealed support for the hypotheses that observation learning is more efficient when few unique exemplars had been experienced, whereas feedback learning is more efficient when many unique exemplars had been experienced in training. The modelfit revealed a dominance of CAM for both observation and intervention. It is possible that the hypothesized processes invited by the learning formats are more easily detected early in the training as in Henriksson et al. (2012). However, the relative fit of the models and the accuracy of judging old and the matched new exemplars was in line with the predictions and suggested that exemplar memory is invited by observation learning and cue-abstraction is invited by feedback learning. The result is in line with the results reported by Rouder and Ratcliff (2004) suggesting that a rule-based process is better able to exploit a task environment when there are many exemplars and that exemplar memory is better able to exploit a task environment when there are few exemplars.

The results in this paper opens for the possibility that the “rule bias” in many studies on categorization and multiple-cue judgment (e.g., Ashby et al., 1998; Erickson & Kruschke, 1998; Juslin et al., 2008; Kalish et al., 2005) may in part be reinforced by the frequent use of the feedback learning format in experiments. However, with a different learning format such as observation there might have been a “bias for exemplar memory” instead. As Juslin et al. (2008) suggest exemplar processes might act as a back-up process that are used whenever rule-based processes fails.

Ashby et al. (2002) has suggested that observation learning might be a learning format that captures many learning situations for children, as when parents teach their children by pointing to objects or persons in the environment and the child is assumed to learn by observing the characteristics of the object. The results from this experiment are in line with previous research that observation is associated with more exemplar processes (Estes, 1994; Henriksson et al., 2012). There is some evidence that 9 to 11 years olds compared to adults have difficulties using cue-abstraction and instead rely on exemplar processes even when a task environment facilitates cue-abstraction. Not fully matured frontal lobe structures, important for working memory, is one explanation for the observed difficulties in using cue-abstraction among the preteen children (Von Helversen, Mata, & Olsson, 2010). Aging might also affect working memory capacity, and as has been shown in categorization, younger adults and elderly perform at similar levels when learning is based on observation learning. But when learning is based on feedback, younger adults outperform older, suggesting that working memory and set-shifting abilities are important in feedback learning (Schmitt-Eliassen et al., 2007). One successful application of the idea of different learning formats is that observation learning seems to offer patients with Parkinson’s disease a way to learn that circumvents their deficits for rule-based processing in categorization (Shohamy et al., 2004). The result from my study suggest that you also can tailor fit the training with few or many exemplars to enhance judgment accuracy by changing the learning format.

In this study, two generic or archetypical cognitive processes in their pure form have been compared, but it is of course possible that people rely on a mix of processes. It is possible that the invited processes in different learning formats can in combination with demands from the task environment transform into a hybrid process and in the future it is reasonable to incorporate models such as SUSTAIN (Love, Medin & Gureckis, 2004) or the Varying Abstraction Model (Vanpaemel & Storms, 2008) to name a few. In terms of such mixed or hybrid models, the results reported here can be understood as a change in the relative dominance of the two processes, where observation learning invites relatively more EBM and feedback learning invites relatively more CAM.

References

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

Ashby, G., Maddox, T. ,& Bohil, C. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition, 30,* 666-677.

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.

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

Estes, W. K. (1994). *Classification and Cognition.* New York: Oxford University Press.

Henriksson, M. P., Enkvist, T., & Juslin, P. (2012). Bias for rules: A question of learning format and task environment. Manuscript submitted for publication.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple-cue judgment: A division-of-labor hypothesis. *Cognition, 10*6, 259-298.

Kalish, M. L., Lewandowsky, S., & Davies, M. (2005). Error driven knowledge restructuring in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 846- 861.

Karlsson, L., Juslin, P., & Olsson, H. (2008). Exemplar-based inference in multi-attribute decision making: Contingent, not automatic, strategy shifts? *Judgment and Decision Making 3*, 244–260.

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

Love, B., Medin, D., & Gureckis, T. (2004). SUSTAIN: A network model category learning. *Psychological Review*, *11*, 309-332.

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

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General*, 115, 39-57.

Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. Journal *of Experimental Psycholigy: General, 133*, 63-62.

Shohamy, D., Myers, C. E., Grossman, S., Sage, J., Gluck, M.A., & Poldrack, R.A. (2004). Cortico-striatal contributions to feedback-based learning: Converging data from neuroimaging and neuropsychology. *Brain, 127*, 851- 859.

Schmitt-Eliassen, J., Ferstl, R., Wiesner, C., Deuschl, G., & Witt, K. (2007). Feedback-based versus observational classification learning in healthy aging and Parkinson's disease. *Brain Research, 1142*, 178-188.

Vanpaemel, W., & Storms, G. (2008). In search of abstraction: The varying abstraction model of categorization. *Psychonomic Bulletin & Review*, *15*, 732-749.

Von Helversen, B., Mata, R., & Olsson, H. (2010). Do children profit from looking beyond looks? From similarity-based to cue-abstraction in multiple cue judgment. *Developmental Psychology, 46*, 220- 229.