Title
Algorithm, Heuristic or Exemplar: Process and Representation in Multiple-Cue Judgment

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 22(22)

Authors
Jones, Sari
Juslin, Peter
Olsson, Henrik
et al.

Publication Date
2000

Peer reviewed

Algorithm, Heuristic or Exemplar: Process and Representation in Multiple-Cue Judgment

Sari Jones (Sari.Jones@psyk.uu.se)
Department of Psychology, Uppsala University
Box 1225, SE-751 42, Uppsala, Sweden

Peter Juslin (Peter.Juslin@psy.umu.se)
Department of Psychology, Umeå University
SE-901 87, Umeå, Sweden

Henrik Olsson (Henrik.Olsson@psy.umu.se)
Department of Psychology, Umeå University
SE-901 87, Umeå, Sweden

Anders Winman (Anders.Winman@psyk.uu.se)
Department of Psychology, Uppsala University
Box 1225, SE-751 42, Uppsala, Sweden

Abstract

We present an experimental design that allows us to investigate the representations and processes used in human multiple-cue judgment. We compare three ideal models of how knowledge is stored and applied in a judgment: A linear additive model (LAM), a heuristic model, Take-the-best (TTB) and a generic exemplar-based model (EBM). The results show that people adaptively change processing depending on what information is present in the learning phase and whether or not the learning situation is compatible with the test. Feedback on a continuous variable provides information sufficient to estimate a LAM that can be used both when learning is and is not compatible with the test. When only dichotomous feedback is provided, the processes differ depending on the learning-test compatibility. At high compatibility, the processing is best described by EBM, but at low compatibility heuristic processes such as TTB become more frequent alternatives to LAM.

Introduction

In the 1950's and 60's two new research paradigms emerged in cognitive science, categorization research (e.g., Shepard, Hovland, & Jenkins, 1961) and research on multiple-cue judgment (e.g., Hammond, 1955). While the former has continued to flourish, the Brunswikian inspired judgment research quietly left the arena in the 80's, although with a re-emergence in studies on realism of confidence (Gigerenzer, Kleinbölting, & Hoffrage, 1991; Juslin, 1994). The two paradigms have a lot in common, but there is seldom cross-reference between them (but see Kruschke & Johansen, 1999). A major conclusion from research on multiple-cue judgment is that linear models fit judgment data well (Brehmer, 1994) but in regard to knowledge representation and processes, there has been little research. In the categorization literature in contrast, a variety of models with explicit representational and process assumptions have been proposed (see Medin, 1989).

In this article, we bring the two paradigms together by combining multiple-cue learning with theories and methods from research on categorization. We present an experimental design that allows us to investigate knowledge representations and processes in human judgment. As a point of departure we take research that postulates multiple levels of representation (e.g., exemplars, rules) that compete to control the judgments in a specific task (Ashby et al., 1998; Erickson & Kruschke, 1998; Logan, 1988). The idea is that experience with some domain may lead to co-existing representations at several levels. A general hypothesis is that the process and representation that dominates at the time of judgment is contingent on an interaction between the learning environment and the judgment task to which the knowledge is later applied. We offer some preliminary ideas in regard to the principles that determine this interaction. The crucial questions is: In what circumstances will a particular level of representation dominate the judgments?

We compare three ideal models of how knowledge is stored and applied in a multiple-cue judgment task. *Linear Additive Models* (LAM) suggest that we store information in memory about: (a) the weight or *cue validity* attached to each cue in the form of a linear coefficient, and (b) an *algebraic rule* for the combination of the cues, in this case a linear additive rule (Brehmer, 1994). The process at the time of judgment is *cue-integration*. Recently, a simpler and more heuristic alternative has been proposed in terms of the *Take The Best* algorithm (TTB; Gigerenzer & Goldstein, 1996; Gigerenzer, Todd, & ABC Group, 1999). TTB suggests that the single most valid cue that is applicable is used and that no information is integrated. The knowledge in memory are *cue validities* and the process amounts to *cue-substitution*. Finally, *Exemplar-Based Models* (EBM) from the categorization literature (e.g., Medin & Schaffer, 1978; Nosofsky & Palmeri, 1997) assert that the memory traces of each encountered object are stored in memory and that judgments are based on the similarity between the new object and the already stored exemplars. In this case, the representations are *exemplars* and the process is similarity-based *retrieval* from memory.

Overview of the Experimental Design

The design presented in this article is based on the idea that an object is judged according to cues. The participants learn that there is a species of frogs that vary in degree of toxicity (0 to 100 %). This attribute depends on four characteristics of the frog; color of the back (green or brown), shape of a spot on the back (wedge shaped or round), size of glands above the eyes (large or small) and color of the abdomen (white or light yellow). These characteristics are binary cues that have the weights .4, .3, .2 and .1 respectively, in a linear equation: Toxicity = $.4 \times$ Cue 1 + $.3 \times$ Cue 2 + $.2 \times$ Cue 3 + $.1 \times$ Cue 4. The weights can be understood as the proportions of poison that each cue adds to the total amount of poison.

Table 1: The exemplars and total proportion of poison when the weights on Cues 1 to 4 are .4, .3, .2, and .1 respectively.

Exemplar	Cue 1	Cue 2	Cue 3	Cue 4	Total
1	1	1	1	1	1
2	1	1	1	0	0.9
3	1	1	0	1	0.8
4	1	1	0	0	0.7
5	1	0	1	1	0.7
6	1	0	1	0	0.6
7	1	0	0	1	0.5
8	1	0	0	0	0.4
9	0	1	1	1	0.6
10	0	1	1	0	0.5
11	0	1	0	1	0.4
12	0	1	0	0	0.3
13	0	0	1	1	0.3
14	0	0	1	0	0.2
15	0	0	0	1	0.1
16	0	0	0	0	0.0

Adding the weights of all cues with positive cue values (1) gives the total proportion of poison for each frog. For example, if a frog has a green back (1), round spot (0), large glands (1) and light yellow abdomen (0) the total proportion of poison is $1 \times .4 + 0 \times .3 + 1 \times .2 + 0 \times .1 = .6$. It is convenient to describe the frogs according to their binary code. The above frog is Exemplar 6 (1010) in Table 1. With four cues, there are 16 possible exemplars. The proportion of poison varies between 0.0 and 1.0 (see Table 1).

In the *learning phase* participants learn to judge whether an exemplar is dangerous or not. Exemplars with a proportion of poison above .5 belong to the category Dangerous, whereas exemplars with a proportion of poison below .5 belong to the category Not dangerous. The participants receive *dichotomous feedback* about the accuracy of their prediction (e.g., "Correct" or "Wrong"). In addition they may or may not receive *continuous feedback* about the exact proportion of poison (e.g., "The amount of poison is 70%"). Exemplars that have a proportion of poison of exactly .5 are randomly assigned as dangerous or not. Three exemplars are omitted in training to test EBM, as described below.

The *single-object test* is the same task as in the learning phase, except that also exemplars that were omitted in the training phase are introduced. In the *pair-comparisons test*, two exemplars are compared in regard to degree of toxicity (i.e., which is the most dangerous one). Feedback is withheld in the tests. The two tests allow us to investigate how knowledge acquired in one task is applied to a new judgment task, and how this affects the choice of representation.

Model Specifications

LAM The characteristic of LAM is that cue validities and a linear, additive integration rule are retrieved from memory and all four cues are weighed together to calculate a total proportion of poison for each exemplar. Clearly, LAM provides the optimal algorithm for the task: If participants have accurate estimates of the cue validities (linear coefficients) they will categorize all exemplars in the single-object test correctly, except when the proportion of poison is .5. With a correctly estimated LAM the accuracy is .94 (15 out of 16 judgments correct), which provides a ceiling on the accuracy that can be attained.

In the pair-comparisons test, LAM computes an estimated toxicity of each of the two exemplars and decides on the exemplar with the higher estimated toxicity. With a correctly estimated LAM all judgments are correct, except when comparing exemplars with the same poison proportion, where the judgment is a guess (e.g., when comparing Exemplars 4 and 5). When all exemplars are compared to each other once (120 comparisons) the success rate is .98 (5 out of 120 comparisons will be guesses). The representations—the cue validities and the integration rule—and the process of cue integration are the same for all exemplars. Therefore, in the test phase LAM predicts that there should be no difference in the accuracy for (old) exemplars previously presented in the learning phase and (new) exemplars that have not been encountered previously.

TTB The characteristic of TTB is that only one of the four cues is used to make a judgment: The cue considered to be the most valid one. In the context of TTB, cue validity is defined as the conditional probability of a correct choice across all cases where the cue is applicable. In the single-object test, TTB implies that an exemplar with cue value 1 on the most valid cue is categorized as dangerous, and an exemplar with cue value 0 as not dangerous. In this application, the most valid cue (Cue 1) has validity .81 when applied to all 16 exemplars. The highest accuracy attainable with TTB in the single-object test is therefore .81.

In the pair-comparisons test, a cue is applicable if one exemplar has cue value 1 and the other 0 on this cue. If the best cue is applicable a judgment can be made, but if the cue values are identical for both exemplars the second to best cue has to be considered, and so on. The most valid cue (Cue 1) has a cue validity of .95 in the application to the pair-comparisons test. The second to best cue (Cue 2) has a validity of .93. In the pair-comparisons test the highest possible accuracy with TTB is .95. This presumes consistent application of the most valid cue that is applicable, correct conception of cue-directions, and correct rank ordering of cue validities. Because the representation and process is the same, TTB predicts that there should be no difference between new and old exemplars.

In addition to these global indices, there are critical exemplars for each cue that discriminate TTB from the other models. In the single-object test, there are two critical exemplars for each cue. If TTB is used, both of these exemplars will be judged incorrectly, whereas if LAM is used they will be judged correctly. Exemplars 8 (1000) and 9 (0111) will be incorrectly categorized if Cue 1 is attended to, because the judgment is based on the binary value of Cue 1 instead of the total poison proportion of the exemplar. Similarly, in the pair-comparisons test there are comparisons that signal the use of TTB. There are three critical comparisons that will be judged incorrectly when Cue 1 is used, because the judgment is based on the binary value of Cue 1 rather than the overall poison proportion of the exemplars. When Cue 2 is used as the best cue, 10 comparisons are critical.[1]

EBM With EBM the new exemplar is judged by similarity to retrieved memory traces of previous exemplars. In the single-object test we assume that the process is well-captured by a standard exemplar-based model from the literature, *the context model* (Medin & Schaffer, 1978). On this account, the new exemplar (probe) is assigned to the category Dangerous with a probability equal to the proportion of the summed similarity to the stored exemplars in the category Dangerous, relative to the summed similarity to all stored exemplars. When the context model is applied to the single-object test with parameters that imply extreme specificity (i.e., all similarity parameters equal to 0), the model only retrieves identical exemplars. In this case, the context model produces the same accuracy as LAM, .94.

Application to the pair-comparisons test, where the two objects (probes) are compared in regard to a continuous variable, is more complicated. In a learning environment with only dichotomous feedback, an EBM like the context model can merely differentiate between pair-comparisons that contrast exemplars from different categories. This holds for 50 percent of the comparisons, and for these an EBM can attain perfect accuracy. For the remaining comparisons that involve two dangerous or two non-dangerous exemplars the judgment has to be based on a guess associated with .5 accuracy. Thus: in a condition with only dichotomous feedback an EBM can at most attain an accuracy of .75 ($.5 \times 1 + .5 \times .5$). When feedback is continuous, on the other hand, an EBM can potentially store also the continuous value with the exemplar. An extension of the original context model that applies to estimation of continuous variables is *PROBEX* (for PROBabilities from EXemplars, Juslin & Persson, 1999). When PROBEX is applied to pair comparison it makes one estimate of the continuous variable for each of the two probes. Each estimate is a weighted average of the values on the variable that have been stored with previous exemplars, where the weights are the similarities to the probe. The rule for computation of similarity is the multiplicative similarity rule of the context model. In a pair-comparisons task, PROBEX decides on the probe with the higher estimated value on the continuous variable. Again, if parameters are set so that only identical exemplars are retrieved (weighted), PROBEX allows the same accuracy as LAM, that is, .98.[2] The reader is referred to Medin and Schaffer (1978) and Juslin and Persson (1999) for a complete specification of the models.

For current purposes, it is sufficient to highlight a general property of many exemplar-based models: Because the judgments are based on similarity to stored exemplars and the multiplicative similarity rule implies a particular sensitivity to identical exemplars, accuracy should be higher for old exemplars that correspond exactly to stored exemplars than for new exemplars. It should thus be easier to categorize exemplars, or compare two exemplars, that have been encountered previously, than new exemplars. EBM can be tested by omitting exemplars in the learning phase and later introducing them in the test phase. If EBM is used, the proportion correct should be higher for old exemplars than for new exemplars.

Cost-Benefit Considerations and Learning-Test Compatibility

We will concentrate on two principles that determines the representation and process that dominates in a task: *cost-benefit considerations* and *learning-test compatibility*. As a preliminary step to this analysis, we have to make a few additional assumptions about the three models. We interpret both LAM and TTB to involve conscious, controlled and analytical processes constrained by short-term memory. We expect that short-term memory can hold at most a few elements (e.g., cue validities) active at any moment and that the process requires active mental effort. EBM is memory-based and the retrieval processes are assumed to be preconscious, automatic and to require little mental effort. It is possible, however, that EBM requires a longer period of learning to accumulate a sufficient set of exemplars. We assume that these processes are present simultaneously and compete to control a specific judgment.

The principle of cost-benefit consideration implies that the relative gain of applying a process is weighted against the cost of applying it. In the context of a design like the present one, the gain is accuracy and the cost is mental effort. The cost involved in applying a process concerns both investment in the learning phase (e.g., the effort to estimate linear coefficients with LAM), and in the test phase (e.g., cue integration). Payne, Bettman, and Johnson (1993) have studied cost-benefit considerations in choice of decision rules in multi-attribute decision making. This research suggests that as the cost of applying a mental algorithm (e.g., LAM) increases people adapt and turn to heuristic processes (e.g., TTB). The principle of learning-test compatibility implies that for memory-based processes, the conditions for successful retrieval are optimal when the circumstances at test match those at learning. This principle is supported by an extensive literature on memory, and illustrated by concepts such as the principle of encoding specificity (Thomson

[1] The results for Cues 3 and 4 are not reported because analyses of the data show that, if TTB is used Cue 1 and Cue 2 are most frequently used as the best cue.

[2] The limits on accuracy are conditional on complete knowledge of all 16 exemplars. Omitting 3 exemplars in the learning phase constrains the possibility to estimate linear coefficients and cue validities, and to store exemplars. These deviations are minor and have no effect on the conclusions.

& Tulving, 1970) and transfer-appropriate processing (Morris, Bransford, & Franks, 1977).

We apply these principles to three experimental manipulations: (a) *Feedback quality*: Presentation of continuous or dichotomous feedback in the learning phase. (b) *Test format*: Single-object or pair-comparisons test. (c) *Cue order*: Fixed or varied presentation order of the cues.

Feedback Quality

Presentation of continuous feedback in the learning phase should enhance the use of LAM. Continuous feedback facilitates estimation of linear coefficients for each cue. This decreases the cost required to use LAM and thereby increases its prevalence. Because learning the cue weights with dichotomous feedback is arduous, cost-benefit considerations suggest that participants are likely to resort to a computationally simpler process like EBM or TTB.

Test Format

In single-object tests, learning and test consist of the same task, whereas in pair-comparisons tests the learning phase and the test differ. The principle of learning-test compatibility implies that EBM should be more common in the single-object test, where the conditions at learning and test are identical. This increases the probability of successful retrieval of stored exemplars. Note that LAM and EBM, in principle, allow the same accuracy, but at different costs: LAM allows rapid learning but requires larger mental effort. EBM requires little mental effort but, presumably, more extended training to attain the same level of performance.

The change of context in the pair-comparisons test should increase the rate of responses guided by LAM or TTB, which are not dependent on episodic retrieval. Participants should have less opportunity to rely on memory (EBM), and turn to the analytic processes implied by LAM and TTB. Moreover, when only dichotomous feedback is provided in the learning phase, EBM provides poor guidance in a pair-comparison that concerns continuous values of the exemplars.

The principles of cost-benefit consideration and learning-test compatibility should interact in a specific way. In pair-comparisons, more mental effort is needed to make the judgment compared to in a single-object test. If LAM is used in a single-object test, four cues are weighted and added. In a pair-comparisons tests, the cognitive effort is doubled. This should make participants who use LAM for single-object judgments swap into a less demanding process in the pair-comparisons test. Because of lower training-test compatibility in the pair-comparisons task, however, they are likely to divert to a heuristic algorithm such as TTB rather than to EBM.

Predictions for the first two manipulations are summarized in Table 2. The provision of continuous feedback should allow the participants to estimate a LAM, and this should be particularly evident in the single-object test where application of LAM demands less effort. With dichotomous feedback, EBM should dominate in the single-object test where training and test conditions match, whereas TTB should dominate in the pair-comparisons test where this match is lower.

Table 2: Processes predicted to dominate as a function of the manipulation of feedback quality and type of test.

Test format	**Feedback quality**	
	Continuous	**Dichotomous**
Single-object	LAM	EBM
Pair-comparisons	LAM/TTB	TTB

Cue Order

In the experiment, we presented the cues in fixed or randomly varied presentation-order across trials. Our hypothesis was that a fixed cue order should enhance the use of EBM because this should maximize training-test compatibility. This manipulation produced no effects, an observation to which we return in the discussion.

Method

Participants

Sixty-four persons (41 women and 23 men, mean age = 24.4) participated. All but 4 were undergraduate students at Uppsala University. Participants received a course credit or a cinema voucher worth approximately 75 SEK for participating.

Design and Procedure

Each learning trial consisted of presentation of an exemplar of a fictitious frog species with 4 different attributes (described above) with the information presented in written text. Three exemplars were omitted and the remaining exemplars were judged 10 times each in the learning phase, making a total of 130 trials. For half of the participants, exemplars 4, 9 and 10 were omitted, for the other half exemplars 5, 6 and 7. These exemplars are equal in poison percentage (i.e., 4 is equal to 5, 9 is equal to 6 and 10 to 7). The omission was thus counterbalanced.

The participants answered the question "Is the frog dangerous or not?" and received dichotomous feedback "Correct answer" or "Wrong answer". Half of the participants also received continuous feedback about the percentage of poison of the frog, for example, "70% poison" The weights, .4, .3, .2 and .1, were randomized to different cues for each participant. For half of the participants cues were presented in fixed order (in the same order and spatial location on the list) and for the other half cues were presented in a varied order.

After the learning phase, participants received a single-object test structurally identical to the learning phase, but without feedback. This phase consisted of 16 trials as all exemplars were judged once. Finally, a pair-comparisons test in which 2 exemplars were contrasted, (also without feedback) was administered. The question was: "Which frog is most dangerous?". This test consisted of 120 trials (all exemplars were compared to each other once). Each session lasted 45 min to 1h and 30 min.

Results

Proportion Correct

In the single-object test, the use of LAM and EBM was predicted and the results support these predictions. Table 3 displays the mean proportions correct (M) and 95 % Confidence Intervals (CI) for each condition.

Proportions correct clearly refute the use of TTB in the continuous feedback condition since the confidence intervals do not include .81, the maximum performance possible for TTB. The proportion correct is significantly higher in the continuous feedback condition than in the dichotomous feedback condition, *t(62)= 2.04, (one-tail) p=.03* This is expected if participants, to some extent, rely on LAM but not if they uniformly rely on EBM. In the continuous feedback condition it is easier to estimate cue validities and therefore the proportion correct is expected to be higher in this condition if LAM is used.

Table 3: Proportions correct (M) and 95 % Confidence Intervals (CI) for each condition of the single-object test.

Feedback	**Cue order**		
	Varied	**Fixed**	**Total**
Continuous	M=.86 *	M=.88	M=.87 *
	(CI: .81-.91)	(CI: .80-.88)	(CI: .83-.92)
Dichotomous	M= .82	M= .79	M=.80
	(CI: .78-.87)	(CI: .70-.88)	(CI: .76-.85)
Total	M= .84	M= .84	M=.84
	(CI: .77-.89)	(CI: .80-.87)	(CI: .80-87)

*The CI does not include .81, the maximum performance possible for TTB.

In pair-comparisons, the participants were predicted to use LAM in the continuous feedback condition and TTB in the dichotomous feedback condition. The mean proportion correct is .90 (CI: .87-.94) in the continuous feedback condition and . 80 (CI: .75-.85) in the dichotomous feedback condition. The proportions correct are consistent with TTB. In sum: The proportions correct falsify TTB in the single-object condition with dichotomous feedback, but the proportions correct are compatible with LAM and EBM in all conditions. Presentation order of the cues had no effect on accuracy.

Old and New Exemplars

The differences in proportions correct for old and new exemplars are presented in Figure 1. The only condition in which there is a substantial advantage for old exemplars over new exemplars is the single-object/dichotomous feedback condition, where EBM is predicted to dominate. In the other three conditions in Figure 1, EBM is not supported. When feedback quality is high and the cost of applying a complex algorithm is low, participants can rely on the powerful (but choosy on data) LAM. When feedback is of poorer quality, the cost of applying LAM is too high and the participants resort to the less demanding EBM. This is particularly likely to occur when learning-test compatibility is high (i.e., in the single-object/dichotomous condition).

In Figure 2, proportions correct are displayed separately for new, mixed (comparing 1 new and 1 old exemplar) and old exemplars in the pair-comparison test. In the continuous feedback condition, there is no difference between new, mixed or old exemplars, supporting LAM (TTB is refuted by the proportion correct, see Table 3). More perplexing, in the dichotomous feedback condition there is no difference between old and new exemplars, but the proportion correct on mixed exemplars is significantly lower. This effect reflects a bias to choose the old exemplar, resembling the *recognition principle* discussed in the context of TTB (Gigerenzer & Goldstein, 1996). This principle states that when presented with a pair comparison between two objects, only one of which is recognized, the participants will guess on the recognized object. In sum: In the dichotomous feedback condition, the comparison of old and new exemplars supports EBM in single-object tests and, potentially, TTB in the pair-comparison tests. In the continuous feedback conditions, the results support LAM.

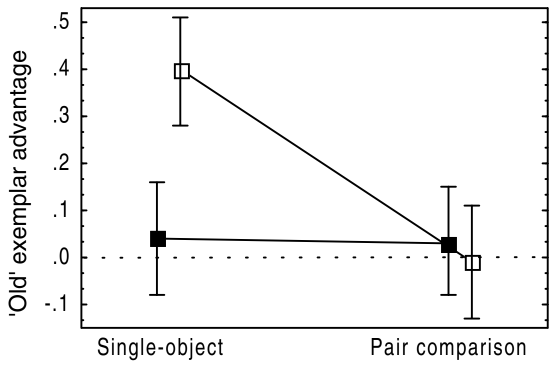

Figure 1: Mean difference between proportion correct for old and new exemplars and 95 % Confidence Intervals for the continuous (filled squares) and dichotomous (open squares) conditions.

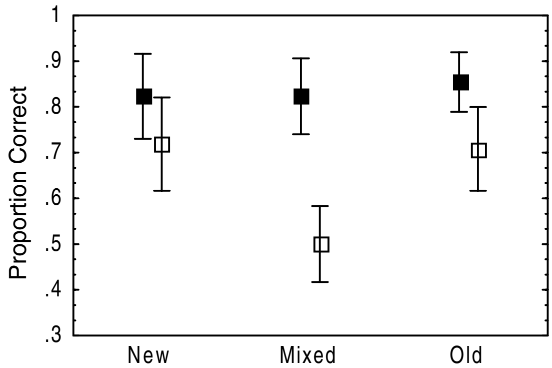

Figure 2: Mean proportion correct and 95 % Confidence Intervals of pair comparisons with new, mixed and old exemplars for the continuous (filled squares) and dichotomous (open squares) feedback conditions.

Critical Exemplars

The critical exemplars do not provide much support for TTB in the single-object conditions. TTB with Cue 1 as the best cue, should yield wrong judgments on Exemplars 8 (1000) and 9 (0111) in the single-object test. The proportion correct on these Exemplars is well above 0 in both conditions, specifically.83 (CI: .72-.94) in the continuous feedback condition and .73 (CI: .62 -.85) in the dichotomous feedback condition. On individual level, five participants (7.8%) used TTB, four judged both critical exemplars of Cue 1 incorrectly and one the critical exemplars of Cue 2. These participants were equally distributed in the dichotomous and continuous feedback conditions. In the pair-comparisons test, TTB receives some support from the critical comparisons. Although a group level analysis shows no consistent use of TTB, on an individual level 16 participants (25%) used TTB: Nine with Cue 1 as the best cue and, surprisingly, seven with Cue 2 as the best cue.

Discussion

We have introduced a design in which the knowledge representations and processes in a multiple-cue learning task can be studied. The results suggest that humans change cognitive processing, as a function of the information present during learning and the compatibility of learning and test, in a way that is consistent with the principles of cost-benefit and learning-test compatibility derived from previous research.

Specifically, the presentation of continuous feedback in the learning phase provided participants with information that allowed them more easily to estimate a LAM. A LAM is applicable both when learning is, and is not, compatible with the test, but the application is more demanding in the pair-comparisons test. Thus, there was more support for the domination of LAM when continuous feedback was provided, but less so in the pair-comparisons than the single-object test.

When only dichotomous feedback is available, the estimation of LAM becomes demanding and other processes come to dominate the judgments. The other processes correspond to the two classical ways of circumventing the limited capacity of controlled thought processes: memory-based performance, or automatization, and heuristic processing. When the test is similar enough to the learning task, processing is memory-based and relies on exemplar representations. When the test is different from learning, heuristic processes, such as TTB, increase in frequency. Overall, however, we found little evidence in support of TTB (i.e., a minority of participants in the pair-comparisons task seemed to rely on it). This may perhaps be explained by the relatively simple task used in the experiment and, indeed, as the task became more complex, evidence in favor of TTB seemed to increase. Nonetheless, at present there is little empirical data that provide support for the empirical validity of TTB.

To our surprise, the manipulation of fixed or varied order of the cues had no effect. Together with the clear effect of old and new exemplars in the dichotomous/single-object condition, this suggests that the judgments are sometimes guided by exemplar-memory, but the representation of the exemplars may be more conceptual than visual in its character.

Acknowledgments

This research was supported by the Swedish Council for Research in Humanities and Social Sciences.

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.

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

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

Gigerenzer, G., & Goldstein, D., G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650-669.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A brunswikian theory of confidence. *Psychological Review, 98,* 506-528.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review, 62,* 255-262.

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57,* 226-246.

Juslin, P., & Persson, M. (1999). *Probabilities from exemplars: On the role of similarity and frequency in probability judgment*. Manuscript submitted for publication.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1083-1119.

Logan, D. G. (1988). Towards an instance theory of automatization. *Psychological Review, 95,* 492-527.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469-1481.

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

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

Payne, J. W., Bettman, J. R., & Johnson E. J. (1993). *The adaptive decision maker.* New York: Cambridge University Press.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75* (13, Whole No. 517).

Thomson, D. M, & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, 86, 255-262.