Multiple-Cue Judgment in Individual and Dyadic Learning

Anna-Carin Olsson (anna-carin.olsson@psy.umu.se)
Department of Psychology, Umeå University
SE-901 87, Umeå, Sweden

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

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

Abstract

Most studies of multiple-cue judgment focus on learning by individuals. In a multiple-cue judgment task we examined if people acquire rule or exemplar knowledge as a function of learning the task alone or in dyads. The expectation was that learning in dyads should promote explicit rule-based thinking as a consequence of increased verbalization (a *social abstraction effect*) and produce a larger joint exemplar knowledge base (an *exemplar pooling effect*). The results suggest more accurate judgments by dyads, an exemplar pooling effect, but no evidence for a social abstraction effect. In contrast to previous research, the social interaction had beneficial effects that allowed participants working in dyads to surpass their combined individual performance.

Introduction

In research on multiple-cue judgment the explicit or implicit cognitive interpretation has often been that people abstract explicit knowledge of cue-criterion relations that is retrieved and mentally integrated into a judgment (Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Juslin, H. Olsson, & A-C. Olsson, in press). Instead of this explicit rule-based knowledge, research on category learning has emphasized that judgments are based on the similarity to memory representations, in particular, to memory traces of *exemplars* (Nosofsky & Johansen, 2000). A developing insight is that a complete account of categorization involves the interplay between multiple qualitatively distinct representations, including both explicit rule-based processes and more implicit memory processes (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Juslin, Jones, H. Olsson, & Winman, 2001; Juslin, et al., in press; Logan, 1988). This mix, or *quasi-rationality*, is also a key-point of cognitive continuum theory of multiple cue judgment (Hammond, 1996).

Almost all research in category learning and multiple-cue judgment involve individuals. These paradigms ignore that judgment is often learned in the context of social exchange, discussion, and collaboration and that judgments often benefit from pooling information across several individuals (but see Ariely et al., 2000, for an exception). While attention is thus paid to multiple representations levels—including non-trivial individual differences in this respect (Juslin et al., in press), the issue of if and how social interaction affects the knowledge representations that a person acquires has not been systematically addressed.

Social Interaction

It has been shown that members of dyads tend to inhibit each other from reaching their maximal memory potential (Andersson & Rönnberg, 1996; Basden, Basden & Henry, 2000). Dyads thus outperform single participants on cognitive tasks, but they do not reach the base-line predicted by the combined performance by the members of the dyad working alone. Explanations involve the *social loafing phenomenon*, stating that a social situation hampers individual productivity because of the lack of personal relevance and motivation (Harkins & Petty, 1982), or *lack of co-operation between group members* (North, Linley, & Hargreaves, 2001). In contrast, the *principle of nonsummativity* (the whole is greater than the sum of its parts: Zaleznik & Moment, 1964) suggests that if a group functions under "psychological independence" the productivity of the group is more than the summed output of the individual members. While the rate of convergence is sensitive to violations of the conditional dyad wise independence, the asymptotic properties are robust under a variety of conditions (Johnson, Budescu & Wallsten, 2001).

The social context influences group communication in different ways (Fleming & Darley, 1991). The increased communication in dyads as compared to individuals working alone suggests a shift to a more analytic representation level, a *social abstraction effect*, because verbal interaction is likely to promote explicit identification of cues and relations between cues. On the other hand, the availability of the exemplars stored by two individuals allows more efficient exploitation of exemplar memory, an *exemplar pooling effect*.

The present study attempts to investigate if people making judgments in dyads develop other processes and knowledge representations than individuals, and, more generally the reasons for differences in performance between dyads and individuals (if such exists).

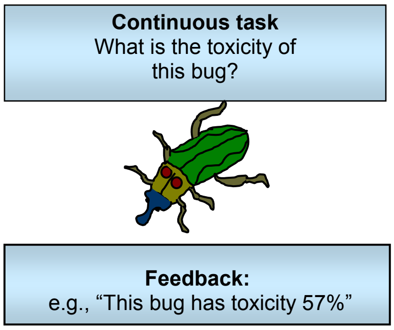

Figure 1: The continuous judgment task.

Judgment Task

The task requires participants to use four binary cues to infer a continuous criterion (Juslin et al., in press). The cover story involves judgments of the toxicity of subspecies of the exotic (but fictitious) Death Bug. The subspecies vary in concentration of poison from 50 to 60 ppm (a continuous criterion), where a concentration below 55 ppm is harmless but a concentration above 55 ppm is lethal. Toxicity can be inferred from four cues of the subspecies (e.g., length of their legs).

The task structure is summarized in Table 1. The binary cues C_1, C_2, C_3, and C_4 take on values 1 or 0. The toxicity c of a subspecies is a linear, additive function of the cue values:

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

C_1 is the most important cue with a *coefficient* of 4 (i.e., a relative weight .4), C_2 is the second to most important cue with a coefficient 3, and so forth. A subspecies with feature vector (0, 0, 0, 0) thus has 50 ppm and is harmless; a subspecies with feature vector (1, 1, 1, 1) has 60 ppm and is dangerous. The 16 subspecies (i.e., possible cue configurations) are summarized in Table 1.

In a *training phase*, participants encounter 11 subspecies. In the Experiment, they make *continuous judgments* about the toxicity of each subspecies (e.g., “The toxicity is 57 ppm”). The judgment task is illustrated in Figure 1. In a test phase, the participants make the same judgments as in the training phase, but for all the 16 subspecies and without feedback. As illustrated in Figure 1, the bugs were presented in an analogue format, as pictures of the cue values.

Cognitive Models

The *cue abstraction model* assumes that participants abstract explicit cue-criterion relations during training that become the objects of mental cue integration at the time of judgment. When presented with a probe, the participants thus retrieve rules connecting cues to the criterion (e.g., “Green back goes with being poisonous”). The rules specify the sign of the relation and the importance of the cue with a cue weight. For example, after training the rule for cue C_1 may specify that C_1=1 goes with a large increase in toxicity.

Table 1: The 16 exemplars with their cues and criteria.

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

Note: E = Extrapolation exemplar, T = training exemplar, O = Old comparison exemplar presented in training, matched on the criterion to one of the new exemplars, N = New comparison exemplar presented the first time at test, p=.5 assigns binary criterion 1 to the exemplar with probability .5.

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

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

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 instructions. Note that Eq. 2 captures the core idea that is crucial to mental cue abstraction: In training, the importance of each cue is abstracted as a cue weight. At the time of judgment, the cue weights are retrieved, applied to the cue values, and integrated into a judgment. *The exemplar model* implies that the participants make judgments by retrieving similar exemplars (subspecies) from memory. When the exemplar model is applied to judgments of a continuous criterion variable, the estimate $\hat{c}_E$ of the criterion c is a weighted average of the criteria c_j stored for the J exemplars, with the similarities $S(p,x_j)$ as the weights:

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

where p is the probe to be judged, x_j is stored exemplar j (j=1…J), $S(p,x_j)$ is the similiarity between probe p and exemplar x_j. Eq. 5 is the *context model* (Medin & Schaffer, 1978) applied to a continuum (see, Delosh et al., 1997; Juslin et al., in press, for similar applications).

The similarity between probe p and exemplar x_j is computed according to the multiplicative similarity rule of the original context model:

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

where d_i is an index that takes value 1 if the cue values on cue dimension i coincide (i.e., both are 0 or both are 1), and s_i if they deviate (i.e., one is 0, the other is 1). s_i are four parameters in the interval [0, 1] that capture the impact of deviating cues values (features) on the overall perceived similarity $S(p,x_j)$. A value of s_i close to 1 implies that a deviating feature on this cue dimension has no impact on the perceived similarity and is considered irrelevant. A value of s_i close to 0 means that the similarity $S(p,x_j)$ is close to 0 if this feature is deviating, assigning crucial importance to the feature. 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, but exemplars with one deviating feature receive weight .001. With s_i close to 1, all exemplars receive the same weight, regardless of the number of deviating features.

Predictions

The predictions are summarized in Figure 2. The models produce similar predictions when the complete set of exemplars are presented both in training and test (the upper panels). Panels A and B illustrate that with appropriate cue weights and low s_i both models predict correct judgments. When the extreme exemplars (*c*= 50 & 60) and three intermediate exemplars (*c*=.55, 56, & 57) are withheld in training, the models produce distinct predictions. In the lower panels the cue abstraction model allows accurate extrapolation beyond the distribution of criteria in the training set [51, 59]. Whenever the correct signs of the cue weights are identified, the most extreme judgments are made for exemplars 1 (*c*=60) and 16 (*c*=50). The exemplar model that computes a weighted average of the observed criteria can never produce a judgment outside the observed range (Delosh et al., 1997). The most extreme judgments are made for criteria *c*=51 and 59.

With the cue abstraction model there should be no systematic difference between judgments for the "New" and "Old" exemplars with *c*=55, 56, and 57: the process is essentially the same in both cases. However, with the exemplar model there is more accurate judgments for Old exemplars: these judgments benefit from retrieval of identical exemplars with the correct criterion.

Effects of Training Alone or in Dyads

In the context of the cue abstraction and exemplar models of concern here, consideration of the task suggested a number of alternative ways in which people can adapt to the demand for learning to make judgments in dyads. A first possibility is an *exemplar pooling effect*. This effect is plausible in a task where the individual participants rely on exemplar memory. When co-operating in dyads they may continue to exploit exemplar memory and together they can store more exemplars in memory leading to improved performance and superior fit for the exemplar model. The exemplar pooling effect comes in a weaker statistical version and a stronger synergetic version. The *statistical version* refers to the mere aggregation effect when the exemplar memories of two individuals are combined. The *synergetic version* implies beneficial effects of working in dyads over and above the mere aggregation effect, for example, because of better encoding of exemplars during training or more efficient retrieval at test.

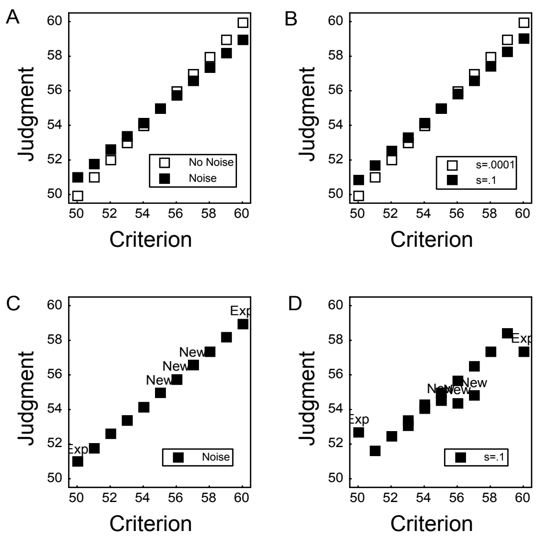

Figure 2: Predictions for the continuous task. Panel A: Cue abstraction models with no noise and noise for the complete training set. Panel B: Exemplar model with all similarity parameter *s* equal to .0001 and .1 for the complete set. Panel C: Cue abstraction model with noise for the constrained set. Panel D: Exemplar model with similarity parameter *s*=.1 for the constrained set.

A second way in which the process may be affected by social interaction is in terms of a *social abstraction effect*. One advantage of abstracting explicit representations of the cue-criterion relations is that it provides knowledge that is more easily communicated by verbal means. While it is relatively easy to verbally explain what specific cues go with high toxicity, it is exceedingly difficult to communicate the entire content of an extensive memory of exemplars. We therefore hypothesized that social interaction, and specifically verbal interchange, should promote a shift from exemplar processes to processes of cue abstraction. The social abstraction effect is evidenced if the data for individual participants is best accounted for with an exemplar model, but the data for dyads is best accounted for with the cue abstraction model.

Method

Participants

Sixty persons participated in the experiment (29 women and 31 men, with an average age of 23.5 years). The participants were undergraduate students at Umeå University and rewarded with 60 SEK with the chance to win 200 SEK. In the experiment the dyads were constructed by participants that already knew each other to lessen the social loafing effect.

Materials and Procedure

The written instructions informed the participants that there were different subspecies of a Death bug. The subspecies differed in toxicity between 50 and 60 ppm. The task was to directly estimate the *toxicity* of the subspecies as a number between 50 and 60. The question on the computer screen was "What is the toxicity of this subspecies". In a training phase the participants received feedback ("This bug has toxicity 57 ppm"). The instructions also informed the participants about the importance of communication to make the judgments in dyads.

The subspecies varied in terms of four binary cues; leg length (short or long), nose length (short or long), spots or no spots on the fore back, and two patterns on the buttock. The cues had the weights 4, 3, 2, and 1 (Eq. 1). The weights determine the portion of toxicity that each cue adds to the total amount. The training phase consisted of 220 trials, where the 11 training exemplars in Table 1 were presented 20 times each. The remaining five exemplars were omitted in the training phase. The four cues were counterbalanced in continuous criteria across the participants.

In the test phase, all participants judged all 16 exemplars, twice. The stimulus formats were presented in two 2x16 blocks, the order of which was counterbalanced across the participants. No feedback was provided in the test phase. All participants were trained and tested with analogue stimuli.

Dependent Measures

Performance is measured by *Root Mean Square Error* (*RMSE*) between judgment and criterion. Model fit is measured by the *coefficient of determination* (r^2) and the *Root Mean Square Deviation* (*RMSD*) between predictions and data in the test phase.

The old-new difference is measured by the difference ΔON_c between the absolute deviations between judgment and criterion (i.e., the absolute judgment error) for the old and the new exemplars with matched criterion *c* (i.e., *c*=55, *c*=.56, and *c*=57 in Table 1):

$$\Delta ON_c = e(c)_{Old} - e(c)_{New}, \quad (7)$$

where $e(c)_{Old}$ is the absolute error for the old exemplar denoted "O" in Table 1 and $e(c)_{New}$ is the corresponding absolute error for the new interpolation exemplar denoted "N" in Table 1. The exemplar model implies that the absolute error is smaller for old exemplars. Cue-abstraction predicts no systematic differences between old and new exemplars. ΔON_c is negative when judgments for old rather than new exemplars are more accurate. For example, if the absolute error from the correct criterion of 57 is 1 for the judgment of the old exemplar (e.g., Exemplar 4 in Table 1) and 2 for the corresponding new exemplar (i.e., Exemplar 5), the ΔON_c is -1 for criterion 57 supporting the exemplar model.

Extrapolation is measured by the observed deviation from linear extrapolation ,

$$Extrap = \begin{cases} (x_{60} - x_{59}) - b, & for \quad x_c = x_{60}, \\ (x_{51} - x_{50}) - b, & for \quad x_c = x_{50}. \end{cases} \quad (5)$$

where (x_{60}-x_{59}) and (x_{50}-x_{51}), respectively, are the slopes of the lines that relate the mean judgments for exemplars with criteria 60 and 59, and 51 and 50 in the additive task. *b* is the difference x_{51}-x_{50} (or equivalently x_{60}-x_{59}) predicted by a linear regression relating mean judgments to criteria (see Figure 1). An *Extrap* of 0 implies that the judgments for the extreme exemplars are as extreme as one would expect from regression-based extrapolation from the old exemplars. *Extrap* is 0 when the judgments are correct and for all linear transformations of the correct judgments. If the index is negative, extreme exemplars do not receive as extreme judgments as expected from extrapolation. For example, if the slope *b* of a regression line when mean judgment is plotted against the criterion is .5, but the slope between x_{59} and x_{60} is -.5 *Extrap* is -1 for the x_{60}-x_{59} comparison supporting the exemplar model with its inability to extrapolate. If the slope for the comparisons x_{60}-x_{59} equals the overall slope *b*=.5 *Extrap* is 0 suggesting appropriate extrapolation.

For ease of exposition of the data reported the two indices are combined into a single exemplar index ΔE for the effects predicted by the exemplar model:

$$\Delta E = \sum_T \Delta ON + Extrap, \quad (6$$

where *T* refers to the overall number of new judgments (both inter- and extrapolation) performed by each participants. In each test block with all 16 exemplars in Table 1, there are five new judgments (3 interpolations and 2 extrapolations). Each participant performed two test blocks so *T* is 10. ΔE is 0 for cue abstraction, but negative in for exemplar memory. (see Juslin, A. C. Olsson & H. Olsson, in press, for further details).

Results

RMSE was lowest for the condition that involved training and test in dyads and highest for training and test as individuals (see Figure 4). *RMSE* was entered into a two-way analysis of variance (ANOVA) with training (individual or dyad) and test (individual or dyad) as between-subjects variables. There were significant effects of both training, (F(1, 38) = 12.52, MSE = .70, p = .00) and test (F(1, 38) = 6.45, MSE =

.70, p = .01), but no significant interaction (F(1, 38) = .00, MSE = .70, p = .98).

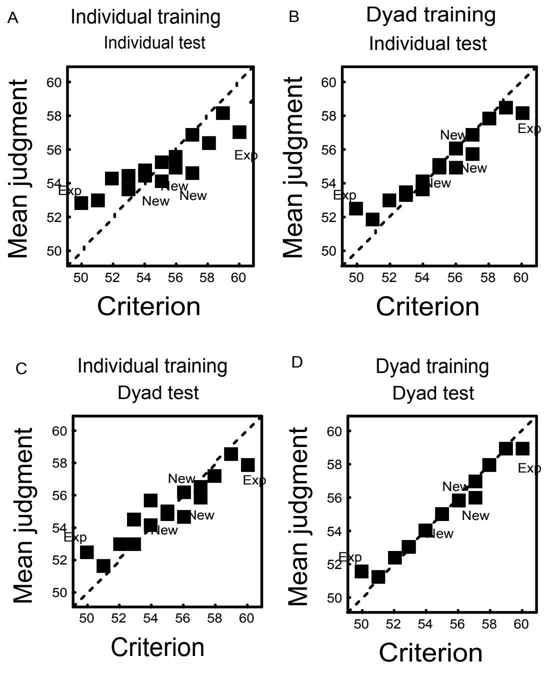

Figure 3. Mean judgments from Experiment 2 plotted as a function of the continuous criterion for the cells with individual training and test (Panel A), dyad-wise training and individual test (Panel B), individual training and dyad-wise test (Panel C), and dyad-wise training and test (Panel D).

The results thus indicate that participants trained and tested in dyads made judgments more accurately than participants trained and tested individually, as illustrated in Figure 3. This beneficial effect arises from additive effects of both training in dyads—supporting the synergetic version of the exemplar pooling effect, as well as of testing in dyads—supporting the aggregation effect implied by the statistical version.

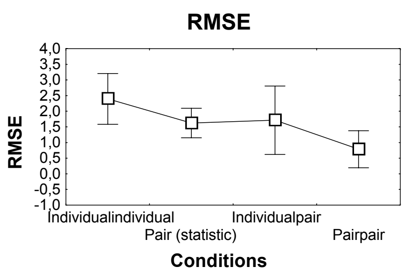

Figure 4. Mean RMSE (Root Mean Square Error) of the judgments with confidence intervals for the individual training-individual test condition, a statistically aggregated dyad condition based on individuals test data, individual training-dyad test condition, and dyad training-dyad test condition.

Figure 4 presents $RMSE$ for the statistical dyads compared to other conditions. The statistical dyads were constructed by forming all possible means based on the judgments by two individuals in the condition with individual training and test. In other words, every judgment is the mean of the judgments made by two individuals that were trained and tested individually. Figure 4 illustrates that the individuals tested in dyads reach the base-line provided by the statistical dyad (i.e., the mean of two participants that have both trained and tested individually), and the individuals that have both trained and been tested in dyads surpass the statistical dyad. These results—together with the significant main effects both of training and testing in dyads—suggest that there is a synergetic exemplar pooling effect.

The exemplar index ΔE was entered into a two-ways analysis of variance (ANOVA) with training (individual or dyad) and test (individual or dyad). There were no significant effects of training (F(1, 38) = .13, MSE = 2.49, p = .72) or test (F(1, 38) = .50, MSE = 2.49, p = .48) and no significant interaction (F(1, 38) = .77, MSE = 2.49, p = .38). The exemplar index ΔE is negative and significantly different from zero in all conditions, with the most negative value in individual training-dyad test condition (-1.08). There is thus no support for the social abstraction effect. The model fits indicate that the exemplar model fits somewhat better for dyad training and the cue abstraction model fits somewhat better for individual training, as illustrated in Figure 5.

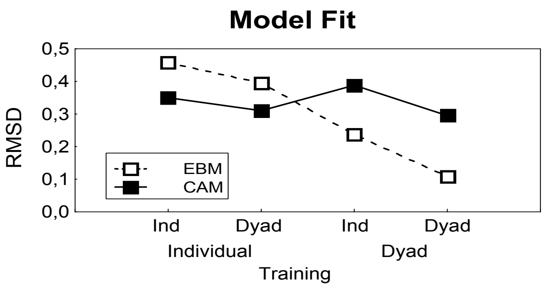

Figure 5. RMSD (Root Mean Square Deviation) between predictions and data for the Exemplar Model (EBM) and the Cue abstraction model (CAM) with individual training and dyad training conditions.

Discussion

The question addressed in this article is if different processes and knowledge representation are developed if people make judgments individually or in dyads and how performance differs between participants working alone or in dyads. We suggested that the effect of social interaction is of importance for understanding our judgments and we wanted to investigate if this factor has an effect on which specific knowledge system that is used in a multiple-cue judgment task. The results showed clear differences in performance between individuals and dyads and indicated that participants that were trained and tested in dyads learned to make judgments more accurately than participants that were trained and tested individually. The ANOVA revealed significant main effects of training and test conditions but no significant interaction, suggesting that the beneficial effect is a simple, additive function both of training and testing in dyads. In Figure 4 the participants trained and tested alone had the highest value of *RMSE*, while the dyads constructed statistically as the mean of two individual judgments have a higher (poorer) *RMSE* than the condition with both dyad training and test. The similar magnitude of the *RMSE* for participants trained in dyads but tested as individuals and the statistical dyads suggest that the individuals tested in pairs were able to reach the base-line predicted by the combination of their individual performance. Because the best fitting model for dyads was the exemplar model (Fig. 5), this illustrates the statistical exemplar pooling effect. The significant main effect of training in dyads suggests additional benefits of training together: a synergetic exemplar pooling effect. An explanation for these results is that in contrast to previous abstract memory tasks (e. g., remembering word lists), this task draws on remembering in the more meaningful context of problem solving. Another potential explanation is that more efficient training allows a shift from cue abstraction to exemplar memory (Logan, 1988).

Performance was different in dyads and individuals, but we were unable to detect any clear differences in the representations or processes. Our results suggest that when co-operating in dyads in a task of the sort addressed here we store more exemplars in memory leading to a more efficient exploitation of memory with exemplar-processes dominating the judgments. Another possibility is that the communication between members of a dyad makes them work with every exemplar more carefully, resulting in a better storage of exemplars.

Acknowledgements

Bank of Sweden Tercentenary Foundation supported this research.

References

Andersson, J. & Rönnberg, J. R. (1996). Collaboration and memory: effects off dyadic retrieval on different memory tasks. *Applied Cognitive Psychology*, 10, 171-181.

Ariely, D., Bender, R. H., Dietz C. B., Gu, H., Wallsten, T. S., & Zauberman, 2001. The effects of averaging subjective probability estimates between and within judges. *Journal of Experimental Psychology: Applied, 6*, 130-147.

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.

Basden, B. H., Basden, D. R., & Henry, S. (2000). Costs and benefits of collaborative remembering. *Applied Cognitive Psychology*, 14, 497-507.

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.

Fleming, J. H. & Darley, J. M. (1991). Mixed messages: The multiple audience problem and strategic communication. *Social Cognition*, 9, 25-46.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducibly uncertainty, inevitable error, unavoidable injustice*. New York: Oxford University Press.

Harkins, S. G., & Petty, P. E. (1982). Effects of task difficulty and task uniquness on social loafing. Journal of Personality and Social Psychology, 43, 1214-1229.

Johnson, T. R., Budescu, D. V., & Wallsten, T. S. (2001). Averaging probability judgments: Monte Carlo analyses of asymptotic diagnostic value. *Journal of Behavioral Decision Making*, 14, 123-140.

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2001). *Cue abstraction and exemplar memory in categorization: Evidence for multiple representation levels*. Manuscript submitted for publication. Department of Psychology, Umeå University, Umeå, Sweden.

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

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

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

North, A. C., Linley, A. & Hargreaves, D. J. (2001). Social loafing in a co-operative classroom task. *Educational Psychology*, 20, 389-392.

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

Zaleznik, A. & Moment, D. (1964). *The Dynamics of Interpersonal Behavior*. New York: Wiley.