Does incidental training increase the prevalence of overall similarity classification?
A re-examination of Kemler Nelson (1984).

Angus B. Inkster (angus.inkster@students.plymouth.ac.uk)
School of Psychology, Plymouth University, Drake Circus
Plymouth, PL4 8AA
Fraser N. Milton (f.n.milton@exeter.ac.uk)
Washington Singer Laboratories, Psychology, College of Life and Environmental Sciences,
University of Exeter, Exeter, EX4 4QG
Andy J. Wills (andy.wills@plymouth.ac.uk)
School of Psychology, Plymouth University, Drake Circus
Plymouth, PL4 8AA

Abstract

Kemler Nelson (1984) reported that incidental training, relative to intentional training, increased the prevalence of overall similarity classification, supporting a non-deliberative account of overall similarity sorting. However, the analysis conducted by Kemler Nelson (1984) does not adequately distinguish between usage of an overall similarity classification strategy and single-attribute strategies. The current study replicates Kemler Nelson's (1984) experiment, seeking to test the original conclusions using a more rigorous analysis. The current study approximates the original experimental procedure, using almost identical stimuli and a longer, modified test phase. Results replicate those found by Kemler Nelson (1984) when the original analysis is applied; however the model-based analysis suggest an overall similarity classification strategy is used rarely and that incidental training increases the prevalence of suboptimal single-attribute strategies. These results imply that overall similarity classification may be more deliberative than previously thought.

Keywords: incidental training, overall similarity classification, single-attribute classification, model-based analysis.

In a seminal study, Brooks (1978) described two processes of categorization. In analytic categorization, the participant separates aspects of the stimulus and evaluates their ability to predict category membership. Brooks assumed analytic categorization would lead to a subset of the stimulus dimensions controlling responding. In non-analytic categorization, the participant predicts category membership on the basis of overall similarity of all stimulus dimensions to known examples. Brooks suggested that non-analytic categorization would be prevalent when a person's cognitive resources were limited. Another way of phrasing this hypothesis is to say that overall similarity classification is considered to be less deliberative than, for example, single-dimension classification. For brevity, we will describe this as the less-is-more hypothesis; e.g. less time spent categorizing objects results in more information from those objects having control over responding. We will contrast this with the more-is-more hypothesis; e.g. more time spent categorizing objects results in more information from those objects having control over responding.

Following the publication of Brooks (1978), experimental evidence seeming to support a less-is-more hypothesis of categorization accumulated. Ward (1983) used a triad procedure, where stimulus triplets are presented, to show that

reducing time for categorization increased overall similarity classifications and decreased dimensionally based classifications. This result was replicated in Smith and Kemler Nelson (1984) who further demonstrated that concurrent load, and instructions that encourage impressionistic responding, increased overall similarity responses. A classic study by Kemler Nelson (1984) used incidental training, where although category membership information is presented on all trials, participants were not directed to attend to it, and compared it to intentional training, where a participant is directed to attend to relevant category membership information to enable correct classification. She found that those who learned the category structure incidentally made more overall similarity classifications than those who learned the category structure intentionally. Smith and Shapiro (1989) followed a similar procedure to Kemler Nelson (1984), but manipulated concurrent load. They found that concurrent load increased overall similarity classifications.

However, not all experimental evidence supported the less-is-more hypothesis. Ward and Scott (1987) found that classifying by overall similarity took longer than by a single-attribute-plus-exception strategy. Several studies found no significant effect of time pressure on overall similarity classifications (Smith & Kemler Nelson, 1984, Experiment 4; Smith & Shapiro, 1989, Experiments 2-3). Nevertheless, the less-is-more hypothesis became widely accepted by the end of the 1990's. Goldstone and Barsalou (1998) stated that, "evidence suggests that in many situations, it is easier for people to base similarity and categorization judgments on more, rather than fewer, properties (p.239-240).

More recent research seemingly provides further support for the less-is-more hypothesis. Waldron and Ashby (2001) demonstrated that concurrent load retarded the acquisition of a single-dimension category structure more than it retarded the acquisition of a multidimensional category structure. Zeithamova and Maddox (2006) used a different stimulus set to Waldron and Ashby (2001), and found a similar effect.

Of course, as with earlier research, not all evidence supported the less-is-more hypothesis. Milton, Longmore, and Wills (2008) reported multiple experiments where increased

time pressure resulted in a reduction of overall similarity classifications in a spontaneous classification task, although one result consistent with a less-is-more strategy was reported. Wills et al. (2009) compared pigeons, squirrels and adult humans under closely matched conditions, but failed to find any species difference in the relative prevalence of overall similarity versus single-dimension classification. Wills, Milton, Longmore, Hester, and Robinson (2013) provide support for a more-is-more hypothesis, reporting three experiments where manipulation of concurrent load reduces overall similarity classification. A further two experiments suggest that those using an overall similarity classification strategy had larger working memory capacities, and that instructions to respond meticulously increase the frequency of overall similarity classifications.

Wills et al. (2013) also provide a critique of the methodologies of the supporting research, in particular the criterial attribute procedure used in Kemler Nelson (1984) and the triad procedure used in Ward (1983). They suggest that none of the methodologies used previously adequately distinguish between less-is-more and more-is-more hypotheses (see Wills et al. (2013) for a more detailed review of these points). The current study focuses upon further investigation of the methodology employed in Kemler Nelson (1984). Kemler Nelson (1984) is widely regarded as a seminal study in the investigation of the less-is-more hypothesis. We refer to their procedure here as the criterial attribute procedure, the abstract structure of which can be seen in Figure 1. There are two methods of attaining 100% accuracy during the procedure's training phase, responding based on the criterial attribute (the dimension of a stimulus that perfectly predicts category membership, D1 in Figure 1), and responding based on information from at least three of the dimensions of a stimulus (e.g. D2,D3 and D4 in Figure 1). The first method is referred to as the Criterial Attribute strategy (a single-attribute strategy), whilst the second method is referred to as the Family Resemblance strategy (an overall similarity strategy).

Training stimuli

| Category A | | | | | Category B | | | |
D1	D2	D3	D4		D1	D2	D3	D4
0	0	0	0		1	1	1	1
0	0	0	1		1	1	1	0
0	0	1	0		1	1	0	1
0	1	0	0		1	0	1	1

| Critical Test Stimuli | | | |
D1	D2	D3	D4
0	1	1	1
1	0	0	0

Figure 1: Training stimuli and critical test stimuli

Participants underwent either incidental or intentional training, followed by a test phase. Participants trained incidentally completed an old/new recognition task, with category membership information about each stimulus incidentally presented, whilst those trained intentionally received a feedback driven categorization task. The critical test stimuli, seen in Figure 1, are used to distinguish between the two strategies, family resemblance and criterial attribute. A person using the criterial attribute strategy would put the first critical test stimulus in Category A and the second in Category B, given that in Figure 1 classifying on the basis of the criterial attribute D1 suggests Category A membership for the first critical test stimulus, and Category B membership for the second critical test stimulus. A person using the family resemblance strategy would respond in the opposite way, as the majority of dimensions for the first critical stimulus suggest Category B membership, and the majority of dimensions for the second critical stimulus suggest Category A membership.

Kemler Nelson (1984) found that adults trained incidentally are more likely to employ a family resemblance strategy than those trained intentionally. Kemler Nelson (1984) is commonly used as evidence in support of the less-is-more hypothesis, but the critical test stimuli used to assess categorization strategy pose a potential problem. Another possible strategy in addition to the family resemblance and criterial attribute strategies is a non-criterial attribute strategy. A non-criterial attribute strategy is another single-attribute strategy, where a participant responds on the basis on a single attribute that does not perfectly predict category membership, e.g. D2 in Figure 1. Use of a non-criterial attribute strategy would result in 75% accuracy during the training phase, enough to pass the learning criterion. The problem arises from the fact that responses that indicate a family resemblance strategy are identical to those indicative of a non-criterial attribute strategy. It is possible that the apparent prevalence of the family resemblance strategy in the incidental condition is partly due to participants using the non-criterial attribute strategy instead. Although Kemler Nelson (1984) does discuss this issue and prevents some fairly informal analyses, we felt the issue merited further, more rigorous, examination.

The present study attempts to determine the strategy used by each participant by looking at responses to all the stimuli presented, as opposed to just the critical test stimuli. The use of a model-based analysis could reveal one of two things, that the family resemblance strategy is indeed prevalent in participants trained incidentally, or that the prevalence of family resemblance strategy found in Kemler Nelson (1984) was due to participants using a non-criterial attribute strategy. With this in mind, the present study aims to replicate Kemler Nelson (1984), using the same procedure but with an extended test phase to facilitate effective usage of the model-based analysis.

Method

Participants and Apparatus

106 undergraduate psychology students at the University of Plymouth participated for course credit. The experiment was conducted on standard PCs, connected to 19 inch flat-screen monitors. Responses were collected using standard PC keyboards. Participants sat roughly 30 cm from the screen. The experiment was run using E-Prime version 2.0.

Stimuli

The stimuli were caricatured line drawings of faces, closely based on those in Experiment 1 of Kemler Nelson (1984), but using ear size as a dimension instead of eye color. Example stimuli are shown in Figure 2. The stimuli varied on four dimensions; hair (straight or curly), ears (large or small), nose (large or small) and moustache(large or small), with the levels of each dimension counterbalanced between participants. One dimension perfectly predicted category membership, with the specific dimension counterbalanced between participants. Each face stimulus was approximately 6cm high and 8cm wide. The 8 stimuli used in training are shown in Figure 1, the test stimuli were all 16 stimuli possible in this four dimensional binary set.
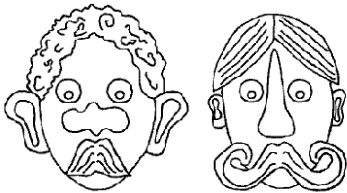

Figure 2: Example stimuli

Procedure

Participants were assigned to one of two conditions, the intentional condition and the incidental condition. The experiment consisted of two sections, a training phase and a test phase. The training phase comprised three blocks of the eight training stimuli seen in Figure 1, with order of presentation randomised within each block. The test phase comprised eight blocks of the sixteen possible stimuli, with order of presentation randomised within each block. In contrast to the original study, all stimuli were presented equally often, whereas in Kemler Nelson (1984) the critical test stimuli were presented more frequently than other stimuli in the test phase. Block breaks in both phases were not signalled to participants.

Training phase. Participants in the intentional condition were told that their task was to identify whether each face was that of a doctor or a policeman. Participants were told that they should try to figure out the rule for identifying whether a face was a policeman or a doctor. Feedback was given for each response, either correcting or confirming their response. Each trial started with the presentation of the two uniforms, to the bottom left and bottom right of the screen respectively.

The uniforms both had a height of 4.5 cm; the doctor's had a width of 5 cm and the policeman's had a width of 9 cm. After 1000ms, the face stimulus appeared in the center of the top of the screen above the uniforms. After a further 2000 ms, "C - Policeman" and "M - Doctor" appeared indicating the response keys to be used. The side of the screen on which each uniform appeared was counterbalanced between participants. Once the participant had responded, feedback was presented for 1500ms. Feedback took the form of the face transposing onto the correct uniform and either the word "correct" or "incorrect" appearing at the top of the screen. After 24 such training trials, participants were moved on to the test phase.

Participants in the incidental condition were told that their task was to decide whether or not they had previously seen a face, before rating their confidence in each decision. No feedback was given on their responses. The faces were presented above the uniform of the group they belonged to; no mention was made of the significance of this. Each trial started with the presentation of the two uniforms, to the bottom left and bottom right of the screen respectively. The side of the screen on which each uniform appeared was counterbalanced between participants. After 1000ms, the face stimulus appeared combined with the correct uniform. After a further 2000 ms the question "Seen before?" and the response keys "Y/N" appeared in the center of the bottom of the screen. After a response was made, the question "Confidence?" and the response keys "1 - 2 - 3" (1 being the least confident, 3 being the most confident) appeared at the center of the bottom of the screen. It is noted that an optimally responding participant in the incidental condition would respond consistently with "N" for each trial after the first block. Whilst this is a somewhat strange task, it does follow the incidental task of the original study. After 24 trials, participants were moved on to the test phase.

Test phase. Participants in the intentional condition were informed that they would be performing the same task, but without feedback. Participants in the incidental condition were told that they should have noticed that in the training phase some of the faces belonged to doctors and some to policemen. They were told that they would see the same faces as in the training phase and that they should aim to categorise them as either doctors or policemen. No feedback was given. The structure of the test trials was identical to the training phase of the intentional condition, without the feedback at the end of each trial.

Results

In Kemler Nelson (1984), participants who failed to reach a criterion of two thirds correct on the training stimuli presented in the test phase were excluded from further analysis. Our test phase was much longer to enable use of the model-based analysis, so the criterion used in the present study was two thirds correct on the training stimuli in the first three blocks of the test phase. This is the closest match to the original criterion that could be used. Applying this criterion left

Table 1: Average number of classification strategy judgements made on critical test stimuli between conditions

Condition	Criterial Attribute	Family Resemblance
Intentional	10.9	5.1
Incidental	7.6	8.4

Table 2: The number of participants best fitting each strategy

Condition	FR	CA	NCA
Intentional	2	19	7
Incidental	1	13	16

29 participants in the intentional conditions and 30 in the incidental condition. In order to achieve these approximately equal group sizes, more participants had to be run in the incidental condition than the intentional condition (35 intentional, 71 incidental). In Kemler Nelson (1984) a similar pattern was noted (17 intentional, 23 incidental, to produce 16 participants in each condition), although the difference was not as extreme as in the present study.

Original study analyses. The first result of interest is the number of criterial attribute judgements and family resemblance judgements made on the critical test stimuli in each condition, which are reported in Table 1. The test stimuli appeared a total of sixteen times in each condition. There were significantly more criterial attribute judgements made on the critical test stimuli in the intentional condition than the incidental condition, $t(55) = 2.14, p = 0.037$. Of mathematical necessity, there are also significantly more family resemblance judgements made on the critical test stimuli in the incidental condition than the intentional condition.

It is also worth noting that participants in the intentional condition made significantly more criterial attribute judgements than family resemblance judgements, $t(28) = 2.48, p = 0.02$. However, participants in the incidental condition did not make significantly more of one type of judgement than the other, $t(29) = 0.37, p = 0.71$.

Model-Based Analysis A model-based analysis was also conducted on the results of the experiment. The analysis consisted of assigning each stimulus the predicted responses of each categorization strategy that we examined (family resemblance, criterial attribute, non-criterial attribute and a response key preference strategy). For each response made in line with the predicted response of a particular categorization strategy, that specific strategy scored a point. This was repeated for each stimulus presented during the 128 trials of the test phase for each individual participant. At the end the scores for each strategy were totalled, and the strategy with the highest consistency score was deemed to be the best fitting strategy for that participant, and the most likely strategy they were employing, a similar method to that employed by Thompson (1994) in the context of a different paradigm. The consistency score represents the number of responses made in line with the predicted responses of a particular categorization strategy. For six of the sixteen different test phase stimuli, the overall similarity strategy can make no prediction (e.g. 1010). For these stimuli the family resemblance strategy scored half a point regardless of the participant's response.

Further Analyses. The number of participants using the family resemblance, criterial attribute and non-criterial attribute strategies are reported in Table 2. One person in the intentional condition was excluded from analysis as response key preference was the best fitting strategy of learning. Inspection of the table reveals two points of interest. First, the family resemblance strategy was rarely used in either condition. Second, incidental training, relative to intentional training, seems to increase the prevalence of non-criterial attribute classification - a sub-optimal single-attribute strategy, $\chi^2(1, N = 55) = 4.5, p = 0.034$.

The average consistency score across the participants in the intentional condition was 116(91%), and in the incidental condition was 103.9(81%), both of which are above chance. The average consistency score for the intentional condition was significantly higher than the average consistency score for the incidental condition, $t(53) = 3.1, p = 0.0031$.

Discussion

The present study found the same effect reported in Kemler Nelson (1984) when applying their analysis, but when the more rigorous model-based analysis was applied found instead that the family resemblance strategy was used rarely and that incidental training, relative to intentional training, increases the prevalence of sub-optimal single-attribute strategies. This does not support the findings of the original study and instead seems to support a more-is-more hypothesis of categorization.

These results seem to contradict those found by Kemler Nelson (1984). A possible explanation lies in the differences between the training received by participants. Intentional training focuses participants on the fact that there is a problem to be solved, and that the trial-by-trial feedback on category membership will help them do this. Under such conditions, many participants presumably work out that one attribute perfectly predicts category membership, and they continue to use this discovery in the test phase.

It is less clear how participants approach the incidental condition. One possibility is that participants, at the beginning of the test phase, select one attribute and respond on that basis (e.g. "policemen have large ears"). The fact that the criterial attribute is selected more often, relative to the non-criterial attributes, than would be expected by chance suggests this selection is not entirely random, but it is presumably harder to identify the criterial attribute from memory than from the online hypothesis-testing permitted by the intentional condition procedure.

One limitation of the current study is that six of the sixteen test stimuli cannot be consistently classified by an overall

similarity strategy (e.g. 1010). This might have lead to participants abandoning the family resemblance strategy upon realising it could not categorize all the stimuli in the test phase, and instead adopting a single-attribute strategy. Despite slightly increasing the number of participants best fit by a family resemblance strategy, the conclusions of the current study are unaffected if those six stimuli are excluded from the analysis, although removing such stimuli from the experiment entirely would be a stronger test of this hypothesis. This work is currently underway in our lab.

The participants in the current study seemed to have a higher failure rate in the incidental condition than in Kemler Nelson (1984). One possible reason for this could be the implementation of the procedure electronically, as completing the procedure 1-on-1 with the experimenter (as was the case in the original study) could have increased participant motivation. Given the large sample sizes required for model-based analyses, individual testing of participants is not a particularly efficient approach. One alternative - currently under investigation in our lab - is to extend the length of training.

One way in which the present study could be further investigated is through the application of formal models such as the widely accepted Generalized Context Model. A recent study by Nosofsky, Denton, Zaki, Murphy-Knudsen, and Unverzagt (2012) investigated incidental prototype-extraction task performance in cognitively impaired patients, applying a modified Elimination By Aspects model to their data to give an explanation of their results. Applying a formal processing model to the present study's data is a work currently being conducted in our lab; we hope to be able to present this at the conference.

Of course, Kemler Nelson (1984) is not the only study to support a less-is-more hypothesis from a criterial attribute procedure. In particular, Smith and Shapiro (1989) is a likely candidate for replication and application of the model-based analysis, given the similarities between the procedures used, and we are currently collecting data for such a replication. Further avenues for possible research include other methodologies for investigating the less-is-more hypothesis, such as the triad procedure employed by Ward (1983) and Smith and Kemler Nelson (1984).

References

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (p. 169-211). Hillsdale, NJ: Erlbaum..

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, *65*, 231-262.

Kemler Nelson, D. G. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior*, *23*(6), 734-759.

Milton, F. N., Longmore, C. A., & Wills, A. J. (2008). Processes of overall similarity sorting in free classification. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 676-692.

Nosofsky, R. M., Denton, S. E., Zaki, S. R., Murphy-Knudsen, A. F., & Unverzagt, F. W. (2012). Studies of implicit prototype extraction in patients with mild cognitive impairment and early alzheimer's disease. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 860-880.

Smith, J. D., & Kemler Nelson, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General*, *113*, 137-159.

Smith, J. D., & Shapiro, J. H. (1989). The occurrence of holistic categorization. *Journal of Memory and Language*, *28*, 386-399.

Thompson, L. A. (1994). Dimensional strategies dominate perceptual classification. *Child Development*, *65*, 1627-1645.

Waldron, E. M., & Ashby, F. G. (2001). he effects of concurrent task interference on category learning: evidence for multiple category learning systems. *Psychonomic Bulletin and Review*, *8*, 168-176.

Ward, T. B. (1983). Response tempo and separable-integral responding: Evidence for an integral-to-separable processing sequence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 103-112.

Ward, T. B., & Scott, J. (1987). Analytic and holistic modes of learning family- resemblance concepts. *Memory and Cognition*, *15*, 42-54.

Wills, A., Lea, S. E. G., Leaver, L. A., Osthaus, B., Ryan, C. M. E., & Suret, M. B. e. a. (2009). A comparative analysis of the categorization of multidimensional stimuli: I. unidimensional classification does not necessarily imply analytic processing; evidence from pigeons (columba livia), squirrels (scurius carolinensis), and humans (homo sapiens). *Journal of Comparative Psychology*, *123*, 391-405.

Wills, A., Milton, F., Longmore, C., Hester, S., & Robinson, J. (2013). Is overall similarity classification less effortful than single-dimension classification? *Quarterly Journal of Experimental Psychology*, *66*, 299-318.

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory and Cognition*, *34*, 387-398.