Better learning of partially diagnostic features leads to less unidimensional categorization in supervised category learning

Sujith Thomas (sujitht@goa.bits-pilani.ac.in)
Department of Computer Science & Information Systems, BITS Pilani, K. K. Birla Goa Campus, India 403726

Narayanan Srinivasan (nsrini@cbcs.ac.in)
Centre of Behavioural and Cognitive Sciences, University of Allahabad, India 211002

Abstract

Previous studies of supervised category learning show that participants often prefer a unidimensional categorization strategy. Studies also report that the perfectly diagnostic feature is learned better compared to the partially diagnostic features. We replicate these results, and we show that better learning of partially diagnostic features leads to less preference for unidimensional categorization. When participants have perfect knowledge about all the diagnostic features, then it becomes equivalent to memorizing the prototypes of the categories. We compare our results with the match-to-standards procedure, where category prototypes are shown during categorization and unidimensional strategy is seldom preferred. We interpret our results to suggest that the preference for unidimensional categorization in supervised category learning, shown in earlier studies, could be due to poor learning of the partially diagnostic features.

Keywords: supervised category learning; observational and feedback learning; unidimensional categorization; memorization of partially diagnostic features; match-to-standards procedure

Introduction

In supervised learning, researchers study the nature of representations that humans use to represent and generalize artificial categories. The supervised category learning paradigm typically has a training phase and a transfer phase. In the training phase, participants learn about the correct category of the training stimuli. The category label associated with each training stimulus depends on the category structure used in the experiment. The training phase in supervised category learning is usually followed by a transfer phase, where participants categorize previously unseen (transfer) stimuli. Usually, all the stimuli are presented serially in both training and transfer phases. Researchers draw inferences about human category learning behaviour based on how the participants categorize the transfer stimuli.

Traditionally two different supervised category learning paradigms have been used, where categories are learned either through observation or through feedback (Estes, 1976; Nelson, 1984). Ashby, Maddox, and Bohil (2002) designed an experiment that compared the observational and feedback learning paradigms. These paradigms differ only in the training phase. In observational learning, participants are shown the category labels along with each training stimulus during the training phase. In feedback learning, participants must categorize each training stimulus, and corrective feedback is provided. It has been shown that feedback learning is more effective than observational learning for category structures that are not based on a unidimensional rule (Ashby et al., 2002).

Levering and Kurtz (2015) have compared classification (feedback) learning and observational learning to see which type of supervised learning leads to better learning of the distributional properties of the features. At the end of the experiment, participants were given a single feature inference test where they judged the category in which a feature is more likely to be in. The single feature inference test would show how well participants could learned the distributional properties of various features. The results showed that participants acquired a perfect knowledge about the perfectly diagnostic feature (i.e. unidimensional criterial attribute), but not about the partially diagnostic features. The results also showed that observational training led to better learning of partially diagnostic features compared to feedback training.

Rabi, Miles, and Minda (2015) used feedback learning to train participants on a category structure that could be correctly learned using either a unidimensional rule (criterial attribute) or by learning the family resemblance structure. In the transfer phase, test stimuli were used that contained the criterial attribute of one category, but shared greater family resemblance with the members of the other category. The study aimed to compare the category generalization pattern among adults and children. We wish to highlight the result that adults preferred to categorize the test stimuli using the unidimensional rule. The second experiment contained a single-feature test phase at the end. In the single-feature test phase, participants were shown all the features one by one, and they had to identify the category in which a feature was most often found. The results showed that the participants — who performed the best in the single-feature test — preferred to categorize the test stimuli using family resemblance sorting strategy instead of the unidimensional strategy (Rabi et al., 2015, P. 164).

The above studies show that in supervised category learning, the (perfectly diagnostic) criterial attribute is learned better compared to the partially diagnostic features. The prototype of a category represents the most common features (i.e. perfectly and partially diagnostic features) of a category (Minda & Smith, 2011). This means that in supervised category learning participants are unable to learn the prototypes of the categories with a high degree of confidence.

Regehr and Brooks (1995) showed that when participants were shown the two prototypes of the underlying categories and asked to categorize the stimuli one by one, participants tend to prefer a family resemblance strategy (and not a unidimensional strategy). This classification procedure has been called match-to-standards procedure. These results have been replicated for stimuli that have distinct and easily identifiable features (Milton & Wills, 2004; Milton, Longmore, & Wills, 2008).

It is not clear whether a unidimensional strategy is always preferred or whether such strategies depend on the learning of partially diagnostic features in category learning. This study aims to address this question. Karpicke and Roediger (2008) showed that it is repeated retrieval (testing) of information that leads to better long term retention compared to repeated studying of information. In our study, we repeatedly tested participants on their knowledge of the partially diagnostic features to achieve better learning of these features. We hypothesized that as participants learn the features in the category prototypes (perfectly and partially diagnostic features) with a greater degree of confidence, there should be a corresponding decrease in the preference for unidimensional categorization. We expected our results to be similar to those of the match-to-standards procedure.

Experiment

In this study, we have made some modifications to the standard supervised category learning paradigm (see Figure 1). Firstly, we use both observational learning as well as feedback learning in the training phase. It has been reported that observational learning leads to (slightly) better learning of partially diagnostic features compared to feedback learning (Levering & Kurtz, 2015). Also, it has been reported that some category structures can be learned better using feedback learning, and not using observational learning (Ashby et al., 2002). Since, our aim was to make participants learn both the perfectly and partially diagnostic features, we have included both observational learning and feedback learning in the training phase of our supervised category learning paradigm.

In the standard supervised learning paradigm, the training phase is followed by the transfer phase. But we have added a memorization phase between the training and the transfer phases as show in Figure 1. In the memorization phase, participants were repeatedly tested on the features that occur more commonly in categories A and B. We had five memorization phase conditions: M0, M1, M3, M4 and M5. In M0 condition, participants memorized task irrelevant features. In M1 condition, participants memorized only the perfectly diagnostic feature, and so on. In M5 condition, participants memorized all five features, that occur commonly in each of the two categories.

In conditions M0 and M1, we expected the perfectly diagnostic feature to be learned with a high accuracy, but not the partially diagnostic features. Also, we expected the unidimensional responses to be high. In condition M5, we expected participants to learn the common partially diagnostic features for the two categories. If the preference for unidimensional categorization is due to poor learning of the partially diagnostic features, then we should expect a significant decrease in unidimensional categorization for condition M5 compared to condition M0.

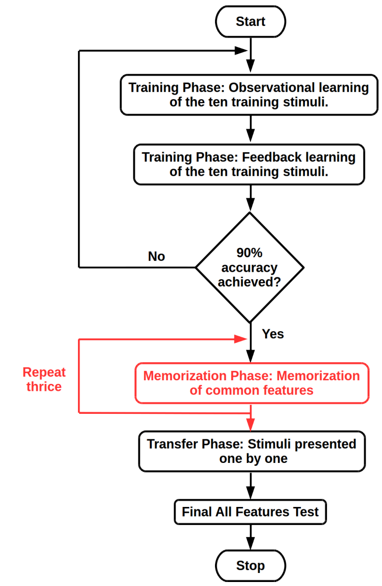

Figure 1: The flowchart shows the experimental procedure. We have added a memorization phase (shown in red) in the supervised category learning paradigm. Also, we use both observational and feedback learning in the training phase. The experimental conditions differ only in the memorization phase.

Method

Subjects We had hundred participants (30 females; mean age = 20.5 years) in this experiment. The experiment used a between-subject design, where participants were randomly assigned to one of the five experimental conditions — M0, M1, M3, M4 and M5. The five conditions had 20 participants each.

Materials Figure 2 shows the fish-like stimuli that were used in Experiment 1. The stimuli have five feature dimensions — shape of the mouth, shape of the upper-fin, shape of the tail, shape of the lower-fin and the body pattern. Each feature along the five dimensions either occurs more frequently in category A or in category B. For example, the pointy mouth occurs more frequently in category A compared to category B. Also, parallelogram-shaped body pattern occurs more frequently in category A compared to category B. One stimulus dimension was perfectly diagnostic of category membership and the remaining four dimensions were partially diagnostic.

The items shown in the first two rows of Figure 2 formed

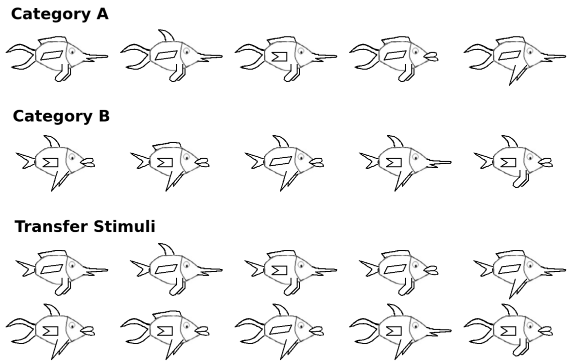

Figure 2: Sample categories used in the experiments. Here, the shape of the tail is the perfectly diagnostic feature, and the remaining features are partially diagnostic.

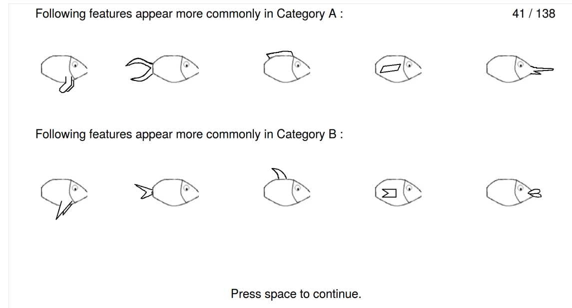

Figure 3: In memorization phase of condition M5, participants were shown all five common features of each category.

the training stimuli, and the items shown in the last two rows formed the transfer stimuli. In Figure 2, a large tail is the perfectly diagnostic feature that occurs in every member of category A. This feature can be called the perfectly diagnostic feature (criterial attribute) because category membership of the training stimuli can be correctly determined using just this one feature. We had five different sets of fish stimuli. A different stimuli dimension formed the perfectly diagnostic feature in each of the five sets. Each set of stimuli was used the same number of times across the five experimental conditions. Figure 2 shows the stimuli in one of the sets, where the tail is the perfectly diagnostic feature.

The transfer stimuli were created by changing the perfectly diagnostic feature of the training stimuli. For example, the transfer stimuli in row three of Figure 2 are same as the category A stimuli in row one, except that the perfectly diagnostic feature (shape of tail) is flipped for the transfer stimuli. Thus, the transfer stimuli are more similar to members of one category, but have the criterial attribute of the opposite category. The category structure that was used in this experiment is same as that used in Rabi et al. (2015).

Procedure Participants were tested in a silent room on a laptop with a 15 inch screen. Participant responses were obtained through the keyboard. Participants had to press A to select category A, and B to select category B. Participants responded to all the trials in a self-paced manner.

The training phase had two parts: observational learning and feedback learning. In observational learning, each of the ten training stimuli were presented, one by one, along with its category label (A or B). The stimuli were presented in a random order. After observational learning participants proceeded to feedback learning. In each trial, participants were presented a training stimulus, which they had to categorize correctly. The training stimuli were again presented, one by one, in a random order. Participants could respond by pressing either A or B. In each trial, participants were given feedback as to whether their categorization response was correct or not. At the end of feedback learning, participants were shown their overall accuracy for the ten training stimuli. Each training phase block consisted of one observational learning sub-block and one feedback learning sub-block. Participants had to achieve 90% accuracy twice (learning criterion) in order to proceed to the memorization phase. If a participant has not achieved 90% accuracy twice, then the training phase was repeated.

In the memorization phase, participants were repeatedly tested about their knowledge of the perfectly and partially diagnostic features of each category; there was no learning criterion that had to be achieved. As discussed earlier, this experiment had five conditions: M0, M1, M3, M4 and M5. These conditions differed in the number of features that participants were made to learn through repeated testing. Participants were not given any feedback.

Condition M5. In the memorization phase of condition M5, participants were shown all the five features that occur more commonly in each category as shown in Figure 3. Then participants were asked to recall the common features of category A. If participants could not recall the features, then they could go to the previous screen and study the common features again. After this, participants were tested about their knowledge of the common features of category A. Participants could click and select the common features of category A. No feedback was given.

The process was repeated for category B. Each memorization phase block consisted of learning the common features of both categories A and B. The memorization phase block was repeated three times. After the memorization phase participants proceeded to the transfer phase.

In the transfer phase, transfer stimuli were presented one by one, and participants could categorize each stimulus into one of the two categories. There were ten transfer stimuli as shown in Figure 2. In the transfer phase, the stimuli were presented in a random order, and no feedback was provided to the participants. The transfer phase was repeated two times for each participant. So there were 20 trials in the transfer phase.

After the transfer phase, there was a surprise all features test phase. In each block, all the 10 features (two features

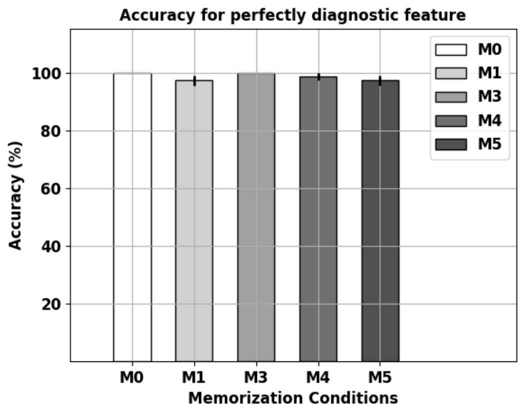

Figure 4: Accuracy for the perfectly diagnostic feature (criterial attribute) in the final all features test.

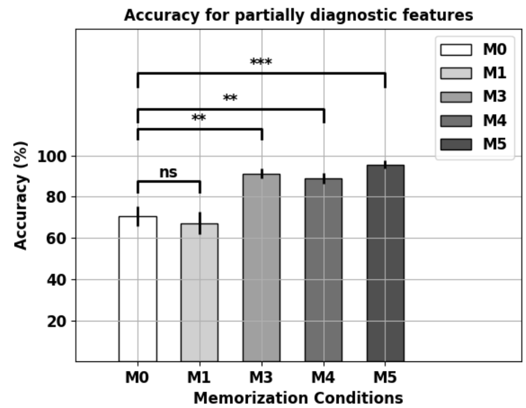

Figure 5: Accuracy for the partially diagnostic features in the final all features test. The significance levels are based on the adjusted p-values found using Tukey HSD test.

along each of the five stimuli dimensions) were presented, one by one, in a random order. Participants were asked to identify the category in which the given feature occurred more commonly. The all features test phase consisted of two blocks, and in each block all the ten features were tested. In the all features test phase, there were 20 trials in total.

Condition M4. Condition M4 was just like condition M5, except that participants learned about only four common features in each category. These included the perfectly diagnostic feature and three partially diagnostic features. All other details were same as that of condition M5.

Condition M3. In condition M3, participants learned about only three common features of each category. These again included the perfectly diagnostic feature and two partially diagnostic features. Other details were same as that of conditions M5 and M4.

Condition M1. In condition M1, participants learned only the perfectly diagnostic feature of each category. So condition M1, should have no effect on the category level knowledge of the partially diagnostic features.

Condition M0. In condition M0, participants were made to memorize features that were not task relevant. These features looked like parts of robot like stimuli, and were very different from parts of the fish stimuli. This condition was included to control for the effect of adding the additional memorization phase in our expermental procedure.

Since participants were not made to memorize any partially diagnostic features in conditions M0 and M1, the results for these conditions should be same as those reported in the earlier studies (Levering & Kurtz, 2015; Rabi et al., 2015).

Results

First we discuss the results of the all features test phase. We expected the accuracy to be high as reported in the previous studies (Levering & Kurtz, 2015; Rabi et al., 2015). Figure 4 shows the accuracy for the perfectly diagnostic feature in the final all features test. The result of one-way between-subjects ANOVA shows that there was no significant effect of memorization on the accuracy of the perfectly diagnostic feature for the five experimental conditions, $F(4,95) = 1.04, p = .39, \eta^2 = .042, \omega^2 = .002$. This shows that participants learned the perfectly diagnostic feature with a high level of confidence irrespective of the experimental conditions.

We expected participants to learn the partially diagnostic features better when their knowledge is repeatedly tested. Figure 5 shows the accuracy for the partially diagnostic features in the final all features test. The result of one-way between-subjects ANOVA shows that there was a significant effect of memorization on the accuracy of the partially diagnostic features across the five experimental conditions, $F(4,95) = 12.06, p < .001, \eta^2 = .34, \omega^2 = .31$.

Post hoc comparisons using the Tukey HSD test (at $p < .05$) indicates that the accuracy for the partially diagnostic features (Figure 5) for condition M0 ($M = 70.63\%, SD = 20.64$) was significantly different from condition M3 ($M = 91.25\%, SD = 10.72$), condition M4 ($M = 89.06\%, SD = 11.33$) and condition M5 ($M = 95.63\%, SD = 8.64$). Also, condition M1 ($M = 67.19\%, SD = 23.87$) was significantly different from condition M3 ($M = 91.25\%, SD = 10.72$), condition M4 ($M = 89.06\%, SD = 11.33$) and condition M5 ($M = 95.63\%, SD = 8.64$). Remaining comparisons were not found to be significant using the Tukey HSD test.

Independent-samples t-test (two-tailed) shows that the accuracy for the partially diagnostic features for condition M0 ($M = 70.63\%, SD = 20.64$) was not significantly different from condition M1 ($M = 67.19\%, SD = 23.87$); $t(38) = .47, p = .64, d = 0.15$. But, the difference in accuracy for the partially diagnostic features between condition M0 and condition M3 ($M = 91.25\%, SD = 10.72$) was significant; $t(38) = 3.87, p < .001, d = 1.22$. Also, the difference in accuracy for the partially diagnostic features between condition M0 and condition M4 ($M = 89.06\%, SD = 11.33$) was significant; $t(38) = 3.41, p = .002, d = 1.08$. Finally,

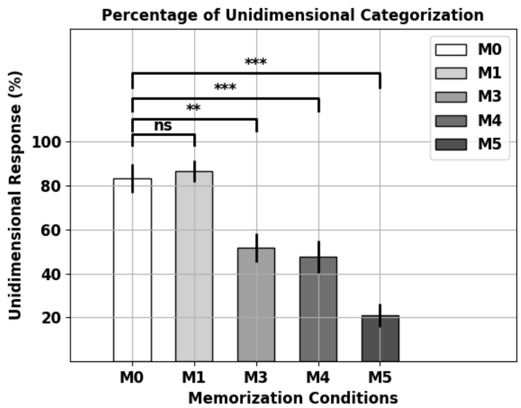

Figure 6: Percentage of unidimensional categorization in the transfer phase. The significance levels are based on the adjusted p-values found using Tukey HSD test.

the difference in accuracy between condition M0 and condition M5 ($M = 95.63\%, SD = 8.64$) was also significant; $t(38) = 4.87, p < .001, d = 1.54$ Overall, the results in Figure 5 show that the category level knowledge about the partially diagnostic features improved when more features were learned through repeated testing.

Studies using supervised learning paradigm report that participants prefer unidimensional categorization (Rabi et al., 2015; Conaway & Kurtz, 2014), and perfectly diagnostic features are learned better compared to partially diagnostic features (Levering & Kurtz, 2015). If poor learning of the partially diagnostic features is leading to the preference for unidimensional categorization, then as partially diagnostic features are learned better there should be a corresponding decrease in unidimensional categorization.

Figure 6 shows the percentage of unidimensional categorization in the transfer phase of the experimental conditions. The result of one-way between-subjects ANOVA shows that there was a significant effect of memorization on the percentage of unidimensional categorization of the transfer stimuli across the five experimental conditions, $F(4,95) = 19.45, p < .001$, $\eta^2 = .45, \omega^2 = .42$.

Post hoc comparisons using the Tukey HSD test (at $p < .05$) indicates that the percentage of unidimensional categorization in the transfer phase (Figure 6) for condition M0 ($M = 83.25\%, SD = 28.25$) was not significantly different from condition M1 ($M = 86.75\%, SD = 21.23$). But, the difference in the percentage of unidimensional categorization for condition M0 was significantly different from condition M3 ($M = 51.75\%, SD = 27.81$), condition M4 ($M = 47.75\%, SD = 32.15$) and condition M5 ($M = 21.0\%, SD = 24.06$).

Independent-samples t-test (two-tailed) shows that the percentage of unidimensional categorization, was not significantly different between condition M0 ($M = 83.25\%, SD = 28.25$) and condition M1 ($M = 86.75\%, SD = 21.23$), $t(38) = .43, p = .67, d = 0.14$. But, results of independent-samples t-test show a significant difference between percentage of unidimensional categorization between conditions M0 and M3 ($M = 51.75\%, SD = 27.81$), $t(38) = 3.46, p = 0.001, d = 1.10$; between M0 and M4 ($M = 47.75\%, SD = 32.15$), $t(38) = 3.62, p < .001, d = 1.14$; and between M0 and M5 ($M = 21.0\%, SD = 24.06$), $t(38) = 7.31, p < .001, d = 2.31$.

Overall, the results in Figure 6 show that as participants learned the partially diagnostic features with a high level of accuracy, there was a corresponding decrease in preference for unidimensional categorization.

Our experimental conditions varied only in the number of diagnostic features that participants were made to memorize. All instructions given to the participants across the five conditions were the same. However, it can be argued that our results are because we have made participants memorize the features and this influenced them to categorize based on the memorized features (experimenter-expectancy effect). Our results would be stronger if, within each experimental condition, we can show a significant correlation between percentage of unidimensional categorization and accuracy for partially diagnostic features.

In condition M5, results of linear regression indicated that there was a significant effect between the accuracy for partially diagnostic features and percentage of unidimensional categorization, ($F(1,18) = 14.12, p = .001, R^2 = .44$, adjusted $R^2 = .41$). The accuracy for partially diagnostic features was a significant predictor in the model ($t = 3.76, p = .001$). In condition M4 also, results of linear regression indicated that there was a significant effect between the two variables, ($F(1,18) = 7.06, p = .02, R^2 = .28$, adjusted $R^2 = .24$). The accuracy for partially diagnostic features was again a significant predictor in the model ($t = 2.66, p = .02$). However, the results of linear regression did not indicate a significant effect between the accuracy for partially diagnostic features and percentage of unidimensional categorization in the following three conditions — condition M3, ($F(1,18) = 2.47, p = .13, R^2 = .12$, adjusted $R^2 = .07$); condition M1, ($F(1,18) = 2.60, p = .12, R^2 = .13$, adjusted $R^2 = .08$) and condition M0, ($F(1,18) = 3.52, p = .08, R^2 = .16$, adjusted $R^2 = .12$).

In conditions M4 and M5, more participants could learn all the partially diagnostic features. So, the effect size is moderate and we obtained a significant effect. For conditions M0, M1 and M3, fewer number of participants could learn all the partially diagnostic features and the effect was not significant even though there was a trend.

As a final analysis, we grouped conditions M0 and M1 together since in both these conditions participants were not made to memorize any partially diagnostic features. Any effect of learning of partially diagnostic features in these groups would indicate that the effect is not solely induced by expectancy created by the task. When conditions M0 and M1 are taken together, results of linear regression indicated that there was a significant effect between the accuracy for partially diagnostic features and percentage of unidimensional categorization, ($F(1,38) = 6.27, p = .02, R^2 = .14$, adjusted $R^2 = .12$). The accuracy for partially diagnostic features was a significant predictor in the model ($t = 2.50, p = .02$). This shows that participants who could learn the partially diagnostic features more accurately were less likely to prefer unidimensional categorization. When only a few participants learn the partially diagnostic features the effect between the accuracy for partially diagnostic features and percentage of unidimensional categorization is low, but the effect becomes moderate when more participants learn all the partially diagnostic features.

Discussion

Multiple studies have investigated the effect of memorization on categorization. Medin, Wattenmaker, and Hampson (1987) investigated the effect of making participants memorize the features of all the items. There were four feature dimensions that took binary values. Participants were made to study and recall the features of the items three times. After memorization, participants categorized the items into two categories. The results showed that participants preferred to categorize using a unidimensional strategy. Similar results were reported by Wattenmaker (1992). Again, participants memorized the features of each item, and then categorized them into two categories. Results show that a majority preferred unidimensional categorization.

In the above mentioned studies (Medin et al., 1987; Wattenmaker, 1992), participants were made to memorize all the features of all the items, but participants still preferred a unidimensional strategy. In these experiments, if participants had changed their strategy, then it could have been argued that the results were due to the attentional bias induced by making participants memorize all the features. The fact that participants continued to prefer a unidimensional strategy was an important finding. In our experiment, we tested whether participants would continue to use a unidimensional strategy after they memorize the common features of each category. It is possible that participants might have continued to prefer a unidimensional strategy due to other reasons like overall similarity based strategy being more effortful (Wills, Milton, Longmore, Hester, & Robinson, 2013). The results of our experiment show that participants do not continue to prefer a unidimensional strategy. Also, we show that in conditions M0 and M1, the preference for unidimensional categorization was negatively correlated with how well partially diagnostic features were learned.

It is possible that after participants memorize the common features they use a more-is-greater strategy, where an item is assigned to a category if it has more features that commonly occur in the category. This change in strategy could have been induced by the memorization phase in conditions M3, M4 and M5. At the same time, if we evaluate studies in which a unidimensional strategy is not preferred, we notice that this occurs when all the items of a category are presented simultaneously in an array (Murphy, Bosch, & Kim, 2017); also, a unidimensional strategy is not preferred for most real world categories (Xu & Tenenbaum, 2007; Abbott, Austerweil, & Griffiths, 2012). In short, a unidimensional strategy is not preferred when participants either know the diagnosticities of different features (e.g. real world categories) or can observe the diagnosticities of different features (e.g. when artifical stimuli are presented in an array). The results in conditions M3, M4 and M5 agree with this pattern, because in these conditions participants know the diagnosticities of multiple features. But to strengthen this argument, a further experiment in which participants learn the prototypes of each category in an implicit manner may be needed.

In our experiment, participants were made to memorize the perfectly and partially diagnostic features of each category. The perfectly and partially diagnostic features are nothing but the features of the category prototypes. So, the M5 condition in our experiment is like the match-to-standards procedure (Regehr & Brooks, 1995); instead of presenting the prototype, we have let participants memorize the prototypes. Our results show that as participants learn the partially diagnostic features with a high level of confidence there is a corresponding decrease in preference for unidimensional categorization. This result resembles the result in Regehr and Brooks (1995), which shows that participants in the match-to-standards procedure did not prefer unidimensional categorization.

Milton and Wills (2004) have shown that the results of match-to-standards procedure are influenced by factors such as features not being distinct and easy-to-identify. Milton et al. (2008) showed that results can also be influenced by time pressure. These results show that match-to-standards procedure may not always lead to a perfect family-resemblance sorting.

Conclusion

It has been shown that in supervised category learning participants often prefer to categorize using a unidimensional strategy (Conaway & Kurtz, 2014; Rabi et al., 2015). Also, perfectly diagnostic feature is learned better compared to the partially diagnostic features (Levering & Kurtz, 2015). Our results suggest that the preference for unidimensional strategy (in supervised category learning) could be due to the fact that partially diagnostic features are not learned with a high level of confidence. There is a lesser preference for a unidimensional categorization strategy once participants learn the partially diagnostic features.

Acknowledgment

Authors acknowledge Department of Science and Technology, Government of India for financial support vide Reference No. SR/CSRI/PDF-49/2016 under Cognitive Science Research Initiative (CSRI) to Dr. Sujith Thomas.

References

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2012). Constructing a hypothesis space from the web for large-scale bayesian word learning. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 54–59). Austin, TX: Cognitive Science Society.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & cognition*, *30*(5), 666–677.

Conaway, N., & Kurtz, K. J. (2014). Now you know it, now you don't: Asking the right question about category knowledge. In *Proceedings of the 36th annual meeting of the Cognitive Science Society* (pp. 2062–2067).

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, *83*(1), 37–64.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *science*, *319*(5865), 966–968.

Levering, K. R., & Kurtz, K. J. (2015). Observation versus classification in supervised category learning. *Memory & cognition*, *43*(2), 266–282.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive psychology*, *19*(2), 242–279.

Milton, F., Longmore, C. A., & Wills, A. (2008). Processes of overall similarity sorting in free classification. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 676.

Milton, F., & Wills, A. (2004). The influence of stimulus properties on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 407.

Minda, J. P., & Smith, J. D. (2011). Prototype models of categorization: Basic formulation, predictions, and limitations. *Formal approaches in categorization*, 40–64.

Murphy, G. L., Bosch, D. A., & Kim, S. (2017). Do americans have a preference for rule-based classification? *Cognitive science*, *41*(8), 2026–2052.

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

Rabi, R., Miles, S. J., & Minda, J. P. (2015). Learning categories via rules and similarity: Comparing adults and children. *Journal of experimental child psychology*, *131*, 149–169.

Regehr, G., & Brooks, L. R. (1995). Category organization in free classification: The organizing effect of an array of stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 347.

Wattenmaker, W. D. (1992). Relational properties and memory-based category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1125.

Wills, A. J., Milton, F., Longmore, C. A., Hester, S., & Robinson, J. (2013). Is overall similarity classification less effortful than single-dimension classification? *The Quarterly Journal of Experimental Psychology*, *66*(2), 299–318.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological review*, *114*(2), 245–272.