COGSCI 2018

changing / minds

40th annual cognitive science society meeting
madison wisconsin usa july 25-28

learning.development.evolution.emergence.plasticity.aging.disorder.cross-culture.experience.persuasion.propaganda.discovery.context.control
emergence.plasticity.aging.disorder.cross-culture.experience.persuasion.propaganda.discovery.context.control.learning.development.evolution
disorder.cross-culture.experience.persuasion.propaganda.discovery.context.control.learning.development.evolution.emergence.plasticity.aging

KEYNOTE SPEAKERS

Susan Gelman
University of Michigan

Matthew Botvinick
DeepMind

Michael Kearns
University of Pennsylvania

INVITED SYMPOSIA

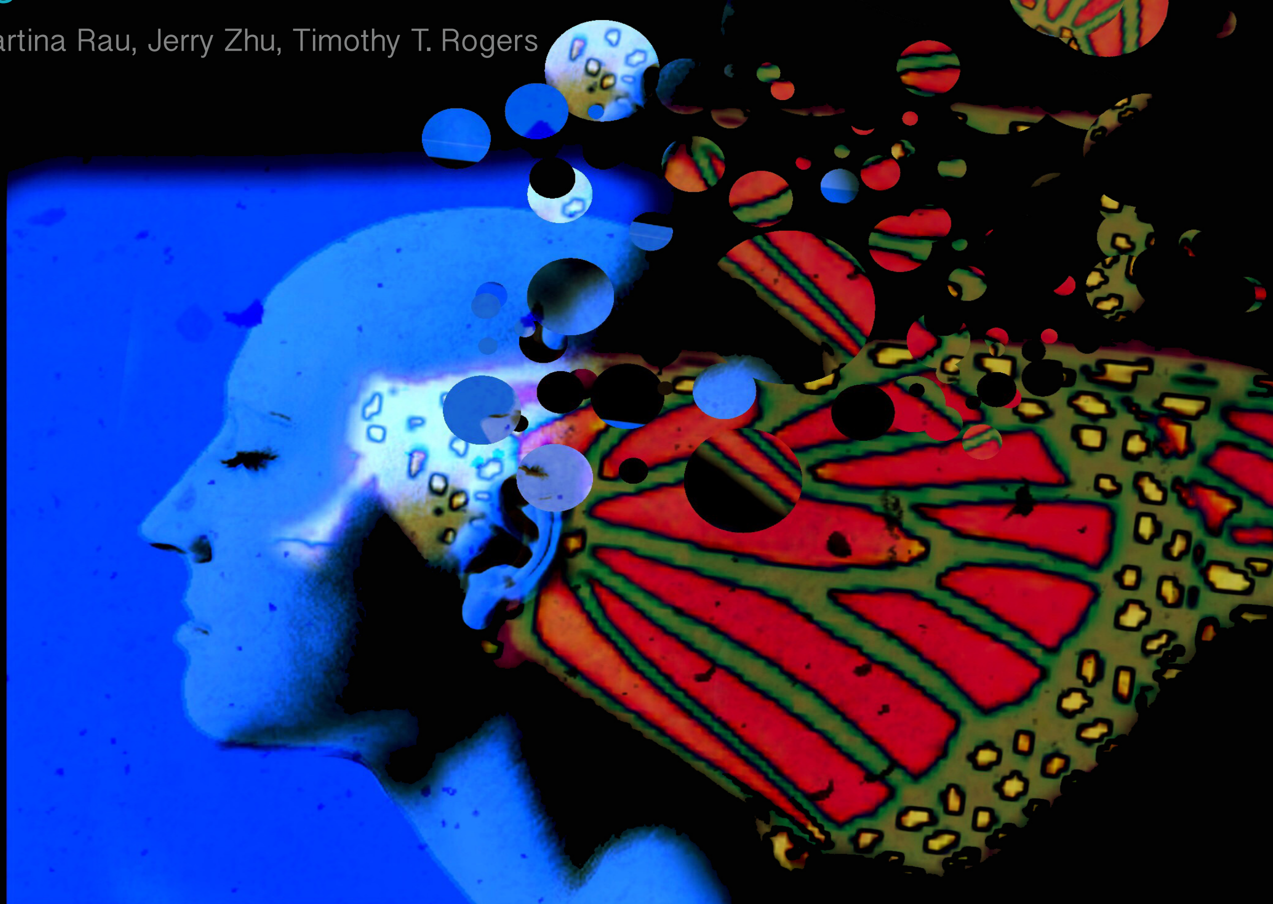
changing education
big data goes to school

changing society
persuasion, propaganda, politics

changing science
Rumelhart prize symposium

ORGANIZERS

Charles Kalish, Martina Rau, Jerry Zhu, Timothy T. Rogers



Perceptual Learning in Correlation Estimation: The Role of Learning Category Organization

Lucy Cui (lucy.cui@ucla.edu)
Christine M. Massey (cmassey@psych.ucla.edu)
Philip J. Kellman (kellman@cognet.ucla.edu)
Department of Psychology, 405 Hilgard Avenue, Los Angeles, CA 90095 USA

Abstract

Research has shown that estimation of correlation from scatter plots is done poorly by both novices and experts. We tested whether proficiency in correlation estimation could be improved by perceptual learning interventions, in the form of perceptual-adaptive learning modules (PALMs). We also tested learning effects of alternative category structures in perceptual learning. We organized the same set of 252 scatter plot displays either into a PALM that implemented spacing in learning by *shape categories* or one in which the categories were ranges of *correlation strength*. Both PALMs produced markedly reduced errors, and both led trained participants to classify *near transfer* items as accurately as *trained* items. Differences in category organization produced modest effects on learning; there was some indication of more consistent reduction of absolute error when learning categories were organized by shape, whereas average bias of judgments was best reduced by categories organized by different numerical ranges of correlation.

Keywords: perceptual learning; category learning; correlation estimation; scatter plots

Introduction

The need to process patterns and relationships in data has never been more prominent than it is today across many aspects of work, citizenship, and daily living. There is growing interest in promoting data literacy, but we are only beginning to understand some of the learning challenges that are involved in doing so. Correlation is one of the most fundamental data relations used across a great variety of contexts, but estimation of correlation from scatterplots, which is how correlations are typically represented, is generally poor and prone to systematic errors.

A chronic problem is observers' tendency to underestimate (e.g., Lauer & Post, 1989; Meyer & Shinar, 1992; Strahan & Hansen, 1978) rather than overestimate (Meyer, Taieh, & Flascher, 1997). Statistically sophisticated observers are no better at estimating correlation than novices, though they do give higher estimates (Meyer & Shinar, 1992). Research on perception and estimation of correlations from scatter plots suggests the influence of a variety of visual features. People tend to give greater correlation estimates when a scatter plot has a greater density of point clouds, even when the points are the same between graphs and only the scale ranges were manipulated (Boynton, 2000). They also tend to give greater correlation estimates to scatter plots with steeper slopes (Bobko & Karen, 1979). Outliers, heteroscedasticity, and restriction of range also affect people's estimations (Bobko & Karen, 1979; Lauer & Post, 1989).

Problems interpreting scatter plots led Doherty and Anderson (2009) to argue for standardizing the graphical features (e.g., axes, labels) of scatter plots in the field of psychology. Standardization of scatter plots, however, does not solve the problem of inaccurate estimation of correlation. In fact, when people see only standardized scatter plots, there is little opportunity to learn to distinguish relevant and irrelevant features; moreover, lack of exposure to non-standardized scatter plots may intensify observers' perceptual biases.

Perceptual learning – experience-induced changes in the extraction of information – is fundamental to this kind of learning challenge (Gibson, 1969; Kellman & Massey, 2013). Research shows that perceptual learning (PL) can be accelerated by interventions involving many short classification episodes that expose the learner to variation within and between learning categories. As the underlying properties (e.g., features, relations) that drive classifications are discovered, perceptual processes come to extract the relevant features preferentially while other irrelevant information may be inhibited. The preferentially selected information comes to be picked up with lower effort or load and ultimately automatically (Kellman, 2002).

The embodiment of perceptual learning techniques in learning technology can be markedly enhanced by combination with particular adaptive learning procedures in Perceptual Adaptive Learning Modules (PALMs; Kellman, Massey & Son, 2010; Thai, Krasne & Kellman, 2015). PALMs systematically put learners through series of classification trials, each dedicated to a particular perceptual classification, which we call a *learning category*. These learning categories are spaced and interleaved adaptively using the ARTS (Adaptive Response-Time-based Scheduling) algorithm (Mettler & Kellman, 2014; Mettler, Massey, & Kellman, 2016), which uses the learner's accuracy and response time on items within a learning category to assess learning strength and determine the learning category's sequencing priority. Remarkably, the same adaptive learning concepts tend to optimize spacing in both factual and perceptual classification domains (Mettler, Massey & Kellman, 2016), a fact likely explained by a general principle – the "successful effort hypothesis" – that applies across learning domains. The key idea is that the best time for another learning trial for a given category in

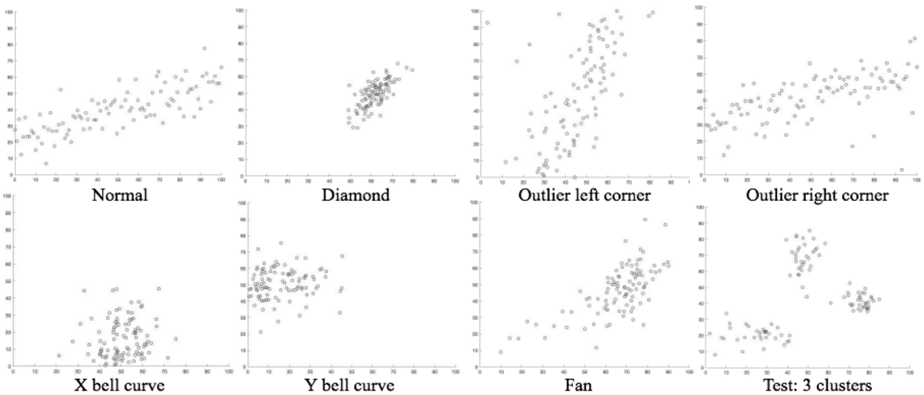


Figure 1. Shapes of scatter plots – manipulated independent from correlation strength. Examples were randomly selected and have different correlations. *Normal* was a linear function with normally distributed errors. *Diamond* was a cluster of dots with its greatest dot density in the middle and least on both ends. *Outlier left corner* was the *normal* shape but with outliers in the left corner, as is *outlier right corner* but with outliers in the right corner. *X bell curve* was a cluster of dots with its greatest density located from the x-axis to center and then tapering off. *Y bell curve* was the same but on the y-axis. *Fan* was a triangular shape with the range of y-values increasing as x increases. *3 clusters* was three *diamond* shaped clusters.

PL is the longest interval at which the learner can still respond correctly (Mettler, Massey & Kellman, 2016).

An unsolved problem of applying perceptual-adaptive learning to category learning is how to determine learning categories. This has usually been done intuitively, where categories are defined such that practice on some instances is likely to advance learning to extract relevant structure in other instances in that category. In some domains, such as diagnostic categories in electrocardiography (Thai, Krasne & Kellman, 2015), the relevant categories are fairly obvious. In other domains, this is not the case. Here we investigate the domain of correlation estimation, in which there are no obvious natural categories, and we test two different schemes of organizing categories spanning the same set of learning instances. We developed two PALMs with learning categories that organized scatter plots based on overall *shape* or *strength of correlation* and manipulated within-category and between-category similarities. The PALMs shared the same scatter plots.

Learning categories in the *Shape* PALM were organized by surface pattern – the shape of the scatter plot. Therefore, they *looked* like "naturally occurring" categories (greater perceptual similarity within categories than between categories), but all learning categories spanned the full correlation range (0 – 1). Thus, there was a large range of correlations within categories that was similar between categories. The learning categories in the *Correlation Strength* PALM were different correlation ranges; within each category, instances could appear with various shapes, and the range of shapes was the same between categories.

The two PALMs provide different learning experiences. Given an incorrect item, the *Shape Category* PALM would set a higher priority for presenting another scatter plot from the same shape category, while the *Correlation Strength* PALM would prioritize another scatter plot with the same correlation range. As a result, participants in the *Correlation Strength* condition might reach mastery of a learning category without seeing instances of that category in every possible shape. Participants in the *Shape Category* condition would continue getting instances of any given learning category until mastery but would not be guaranteed to see all ranges of correlation.

We hypothesized that if one way of organizing learning categories was more compatible with commonalities of perceived structure, that condition might show better learning.

Methods

Participants

103 undergraduates from the University of California, Los Angeles participated for course credit. Four participants were excluded for not completing the experiment.

Materials

We created seven different shapes of scatter plots, inspired by the visual features (e.g., dot density, outliers) that had an influence on correlation estimation in the literature (see Figure 1). The correlation range: 0 – 1 was divided into seven bins (0 – 0.14, 0.15, - 0.28, 0.29 - 0.42, 0.43 – 0.56, 0.57 – 0.70, 0.71 – 0.84, and 0.85 – 0.99). Both PALMs had the same 49 subcategories (from all 7 shapes x 7 correlation range combinations), but subcategories were either arranged into categories based on shape or correlation range. In order to have transfer items at posttest, one correlation range (e.g., 0 - 0.14) was withheld from training for each shape category, with no correlation range omitted more than once across the seven shape categories, and vice versa. The same subcategories were omitted from training for both PALMs.

PALM Parameters Each category was introduced with an initial *passive* trial, a scatter plot displayed with the numerical correlation shown. Each passive trial "unlocked" the learning category and initiated *active* trials, where participants gave a response and received trial-by-trial feedback, for that learning category. In other words, the beginning of training was a combination of *passive* and *active* trials until all learning categories were introduced, then trials were only *active.* If a response was not given within 20 seconds, the trial timed out and the numerical correlation was presented. After every 25 trials (1 block) participants were shown their average accuracy and response time for previous block(s). Participants' correlation estimations were considered correct if they were ±.07 of the actual correlations – chosen to be half the size of a bin (~0.14). Categories were adaptively sequenced using the ARTS sequencing algorithm (see Mettler & Kellman, 2014), which sets priorities for categories reappearing based on participants' accuracy and response time on items in those categories. The minimum number of trials in between items from the same category, or *enforced delay*, was set to two trials. *The enforced delay* parameter in ARTS precludes reappearance of the same category while recent feedback still persists in working memory. Participants "retired" a learning category when they met mastery criteria consisting of 4 correct responses out of the last 5 trials of a category, with RTs ≤ 5 seconds. Training continued until participants graduated from all categories.

Scatter Plots Six unique scatter plots were created for each subcategory used in training (total of 252). The number of data points (100) and the scale of the scatter plot (0 – 100 for x- and y-axes) were kept constant for all scatter plots while slope and intercepts varied. Slope was not correlated with correlation, $r = -0.17$, $p = 0.60$. Because scatter plots look very similar at very low correlation values, scatter plots were considered to have different shapes if they looked distinct at the 0.5 correlation level. Great care was made to ensure equal representation of and no gaps in correlations (i.e., 0.01, 0.02, 0.03, and so on through 0.99 appear at least once in the training set). Variance in x and y values was determined by randomly sampling from a normal distribution with varying means and standard deviations.

Assessments Pretest and posttest items were identical but appeared in different random orders. There were four types of items: 1) *training set*: 7 scatter plots drawn from the training set to represent each feature; 2) *near transfer*: 7 scatter plots from each of the seven subcategories omitted from training, 3) *far transfer*: 7 scatter plots with a novel shape representing each correlation range; 4) *negative*: 4 negative correlation scatter plots spanning -1 to 0. *Far transfer* was considered transfer of correlation estimation skill to a novel shape that is still within the correlation range trained on. The novel shape was selected to have similar visual features (i.e., dot densities) to trained *shapes* but with a dot distribution not seen in training (i.e., three distinct dot clusters). *Negative items* were used to test for remote transfer, as *shapes* and absolute value of correlation range were the same but the data had a negative trend.

Procedure

At the beginning of the PALM, participants were given the definition of correlation and a scatter plot example of a positive, a negative, and no correlation, without actual correlations labeled. They were informed that all scatter plots seen in the experiment would have the same number of data points and the same scaled axes. Participants were asked to give their correlation estimates to the second decimal place and told that their progress through the PALM depended on their speed and accuracy. Before the pretest, participants were informed that scatter plots presented during assessments could have negative or positive correlations. Prior to starting training, participants were reminded that scatter plots during training would only have positive correlations. After training, participants completed an immediate posttest and a survey. Participants were asked to report demographics (age, gender), exposure to statistics (number of courses and average grades), familiarity with the term correlation (heard of it, can define, can interpret, know formula) and strategy for estimating correlation. They also rated their level of frustration, attention, and effort on a Likert-scale from 0 to 5.

Dependent Measures

Performance was measured in several ways. The *absolute deviation* measure reflected the absolute value of the difference between the participants' estimate and the actual correlation. We defined *mean error* as average signed deviation across responses (participant's estimate minus actual correlation). Negative values of *mean error* represented underestimating and positive values overestimating. We defined a binned *accuracy* measure such that an estimate was scored as correct if it fell within ±0.07 of the actual correlation. Because participants learned to mastery criteria, the amount of time spent and number of trials completed during training varied across participants. To account for this, we calculated *learning efficiency scores* by dividing accuracy gain (posttest minus pretest) by minutes or trials.

Results

The primary results of this study are shown in Figure 2. The left panel shows the mean *absolute deviation* of correlation estimates at pretest and posttest, for both *training set* items and *near transfer* items. Both groups showed substantial learning. There were no differences on efficiency measures. There is some indication that the *Shape Category* condition showed more consistent learning for *near transfer* items. The right panel shows *mean* (signed) *error* for estimates, across conditions and tests. Both groups improved from pretest to posttest, with the *Correlation Strength* condition ending up, as a group, with mean posttest estimates not much different from zero. These observations were confirmed by the analyses, described further below.

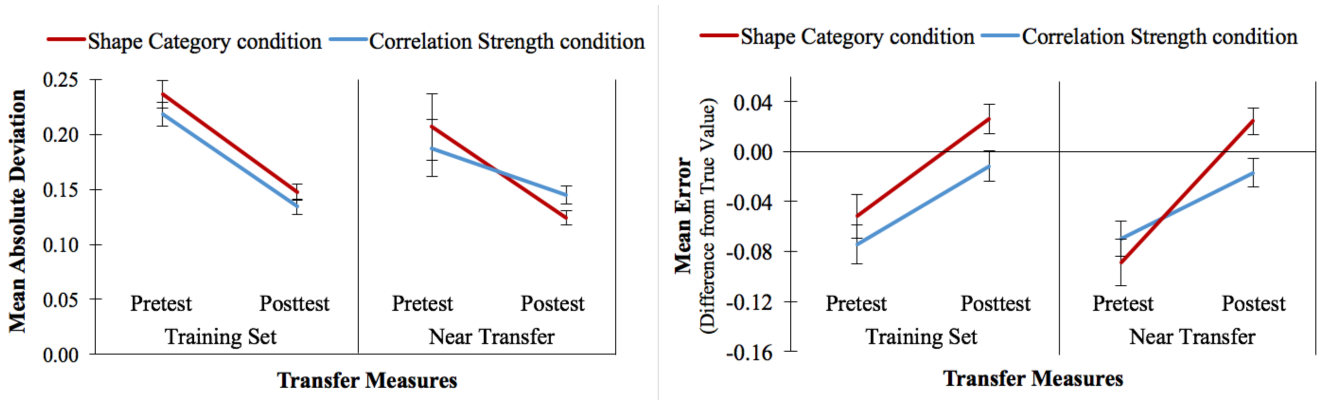


Figure 2. (Left) Improvements in mean *absolute deviation* on *training set* and *near transfer* items from pretest to posttest. (Right) Improvements in *mean error* on *training set* and *near transfer* items from pretest to posttest. Error bars represent +/-1 standard error.

Accuracy Measures

Absolute deviation. We conducted separate one-way ANCOVAs on *absolute deviation* for each of the transfer measures, with the pretest *absolute deviation* for the corresponding transfer measure as a covariate. Posttest scores for the *Shape Category* condition had a lower *absolute deviation* for the *near transfer* items (M = .12, SD = .05) than the *Correlation Strength* condition (M = .14, SD = .06), $F(1,96) = 5.51$, $p = .021$, $\eta_p^2 = .05$. There were no reliable differences for *training set* items or *far transfer* items. For *negative* items, there were no positive learning effects at all; both conditions showed larger *absolute deviation* scores at posttest than at pretest. There was a marginal tendency for greater increases in deviation in the *Shape Category* condition, $F(1,96) = 2.88$, p =.093, $\eta_p^2 = .03$. We suspected that some subjects may have only assessed the strength of correlation, which in the absence of the negative sign, would produce even worse deviation scores than pretest. We calculated the absolute deviation from the magnitude of the participants' responses and of the correct correlations, and ran a 2 condition x 2 assessments (pretest, posttest) repeated measures ANOVA to investigate this possibility. Participants did in fact improve their estimates of strength of correlation, $F(1, 97) = 44.80$, $p < .001$, but there was no difference between conditions, $F(1, 97) = 0.418$, $p = .52$.

An important feature of these results is that *near transfer* items were generally answered as accurately as *training set* items. This result indicates perceptual learning of structural characteristics rather than memorization of instances. There was some suggestion of more consistent improvement by the *Shape Category* condition across *training set* items and *near transfer* items. An ANOVA using condition, test phase, and item type showed a marginally reliable 3-way interaction, $F(1,97) = 3.72$, $p = .057$, $\eta_p^2 = .04$.

Mean Error. *Absolute deviations* sum the absolute values of error (unsigned error) for participant responses. We label *mean error* as the average signed deviation across responses. We conducted separate ANCOVAs on *mean error* for each of the transfer measures, with the pretest mean error for the corresponding transfer measure as a covariate. Positive *mean errors* reflect overestimating and negative *mean errors* reflect underestimating. For the *training set* items, the *Shape Category* condition had a positive mean error (M = *.03, SD = .08*) while the *Correlation Strength* condition had a negative *mean error* (M = -.01, SD = .09), $F(1, 96) = 4.93$, p = .03, $\eta_p^2 = .05$. This pattern also held for the *near transfer* items: the *Shape Category* condition had a positive *mean error* (M = .02, SD = .07) while the *Correlation Strength* condition had a negative *mean error* (M = -.02, SD = .08), $F(1,96)$, p =.008, $\eta_p^2 = .07$. There was no reliable effect of condition for *far transfer* items.

The differences in mean error may indicate different biases across conditions. To assess potential differences in error, we tested all of the pretest and posttest data points for *training set* items and *near transfer* items against the hypothesis of zero error, using one-sample t tests. Both conditions showed mean error reliably greater than 0 at pretest, for both *training set* and *near transfer* items. Posttest results suggested greater improvements in accuracy in the *Correlation Strength* condition than in the *Shape Category* condition. The *Shape Category* condition differed reliably from zero error at posttest for *training set* items $t(46) = 2.17$, p = .04) and for *near transfer* items ($t(46) = 2.31$, p =.03). For the *Correlation Strength* condition, there was no reliable difference from the hypothesis of zero error, either for *training set* items ($t(51) = -.97$, p =.34) or for *near transfer* items ($t(51) = -1.49$, p = .14). In other words, in the *Correlation Strength* condition, participants moved from underestimating to unbiased estimates, while those in the *Shape Category* condition shifted from underestimating to overestimating.

Accuracy Gain. We defined a separate binned accuracy measure such that an estimate was scored as correct if it fell with ±.07 of the actual correlation. We conducted separate ANCOVAs on accuracy gain (posttest minus pretest) for each of the four transfer measures, with the pretest accuracy for the corresponding transfer measure as a covariate. Accuracy improvements were modest. Both conditions improved their accuracy on *training set* items (M = .13, SD =.19) and on *near transfer* items (M = .15, SD = .24) and did not improve their accuracy on *far transfer* items (M = .00, SD = .16) or *negative* items (M = .01, SD = .22), p's > .05.

Usability and Subjective Experience

The PALMs were very similar in length and subjective experiences. There were no differences between conditions in the number of trials completed during training or time to reach learning criterion. Perceiving correlations is difficult. The PALMs were equally frustrating for participants (M = 4.06, SD = 1.07), but participants paid attention just the same (M = 3.48, SD = 0.92), *n.s.* Participants in the *Correlation Strength* condition (M = 3.67, SD = 0.92) reported slightly more effort than those in the *Shape Category* condition (M = 3.34, SD = 0.92), $t(97) = -1.80$, p = 0.08, d = 0.36.

Discussion

Learning technologies designed to improve learning in most domains employ a category structure that mirrors natural categories, such as species for classifying butterflies, diagnoses for reading medical scans, or problem types for practicing mathematics. Sequencing items in this way is intuitive. However, we asked whether alternative category structures could benefit learning and perhaps even yield different learning outcomes. People estimate correlations from scatter plots poorly, even observers seasoned in statistics. We chose correlation estimation to see if we could improve this skill using perceptual learning principles and to see whether different category structures matter.

Some research on correlation estimation suggests that various visual features influence estimation, so for one category structure, we grouped scatter plots by their shapes, whereas we used correlation ranges as an alternative grouping. Both PALMs utilized the same learning items. We predicted that perceptual learning interventions that exposed observers to variation within and between learning categories, involve *active* classification episodes, and provide immediate feedback would increase correlation estimation proficiency in both PALMs but that the degree and nature of improvement might differ between them. We hypothesized that the *Shape* PALM would develop a correlation estimation skill that is more robust with respect to variations in surface features in scatter plots. Another possibility was that participants in the *Correlation Strength* condition would get an advantage in *near transfer* (where some range of correlation had been withheld from the training set), due to getting systematic practice along the dimension of degree of correlation.

We found that both PALMs improved proficiency in correlation estimation - a notable result, as even years of interaction with scatter plots do little to develop experts' ability to extract invariant structure in this domain. Participants did, indeed, train on a substantial number of unique scatter plots (252) and complete many trials (~500 on average) during a condensed time period – a learning experience that is unusual. Although statisticians interact with scatter plots often, they certainly would rarely see this many in succession and certainly not in an order that benefits learning.

Participants also estimated *near transfer* items as accurately as *training set* items. Recall that participants never saw scatter plots with these combinations of shape and correlation range in training. Equivalent performance on these items is consistent with perceptual learning of structural characteristics as opposed to memorizing individual instances.

Although participants in both conditions trained to objective learning criteria, as defined by our accuracy requirement of being ±.07 of the actual correlation, the two PALMs yielded different learning outcomes. Participants in the *Shape Category* condition were more consistent in the amount their estimations deviated from the actual correlation while participants in the *Correlation Strength* condition were less biased in their estimations. In addition, participants in the *Shape Category* condition were significantly closer (lower absolute deviation) to the actual correlation on *near transfer* items than those in the *Correlation Strength* condition.

The reliable differences in bias are not large, but they may reflect differences in learning experiences between the two PALMs. Participants in the *Shape Category* condition, on average, overestimated. Because learning categories in this condition were not systematically organized in terms of degree of correlation, category sequencing based on performance may have been less impactful at addressing bias, despite accuracy feedback. For example, an error on an exemplar from a given shape category with a true correlation of .75 might have been followed up soon after by another example of that shape category, but the new instance could have a very different degree of correlation. In contrast, in the *Correlation Strength* condition, an error on a display with correlation of .75 would be followed up within a couple of learning trials with another category exemplar with a correlation close to .75. Such effects of category structuring might also occur with regard to attainment of learning criteria. A persistent error relating to a given correlation range would tend to delay mastery in the *Correlation Strength* condition, leading to more learning trials centered on that category. A final possible contributor to the condition difference for bias is that overestimating seems to be reflective of statistical sophistication (Meyer & Shinar, 1992). We do not know why, but our data suggest a growth of skill in both conditions, whereas only in the *Correlation Strength* condition would category structure have tended to drive adaptive learning events that might tend to combat consistent overestimation, especially one centered in certain parts of the range of correlations. Perhaps some explanation along these lines explains why the *Correlation Strength* condition showed posttest results for both *training set* items and *near transfer* items that did not differ reliably from zero error.

Conversely, participants in the *Shape Category* condition did outperform those in the *Correlation Strength* condition on *near transfer* items, in terms of absolute deviation. Their improvements on *near transfer* items were consistent with the amount they improved on *training set* items. This difference may speak to a superior pick up of structure and decreased attention to surface features as a result of training, allowing these participants to estimate correlation across a broader range of shapes.

Participants were not able to transfer their correlation estimation skill to a novel shape, as they performed just as poorly on *far transfer* items at posttest as pretest. We suspect that our *far transfer* items were so difficult that transfer would have been close to impossible. Although participants had experienced scatter plots with dot densities during training, they had only one cluster, not three clusters as in the *far transfer* items. Performance on *negative* items got worse after training, which can be explained by the absence of the negative sign in their estimations. When only strength of correlation was assessed, participants did in fact improve from pretest to posttest. Participants may have omitted the negative sign because they became less attuned to slope, as slope varied throughout training and did not correlate with correlations of scatterplots, so noticing this feature was useless and therefore, disregarded. Negative, shallow slopes would be harder to detect at posttest, possibly leading participants to misclassify them as positive, resulting in larger deviations. The same filtering out of surface features that gave participants in the *Shape Category* condition an advantage on *near transfer* items could be a disadvantage when processing surface features becomes relevant to the task (i.e., looking at which way the points are pointing when slope is shallow), as in *negative* items.

To our knowledge, little work has compared different ways of organizing learning categories in complex perceptual learning. The results of this study demonstrate that the perceptual learning intervention was successful in improving novices' skill in the difficult and error-prone task of estimating correlations from scatter plots, and that variations in how the learning categories were defined and sequenced differentially showed some measurable effects on absolute accuracy and bias in estimation. Our results suggest that the type of learning outcome may depend on how learning categories are organized and should be considered when designing learning modules.

The role of learning category organization deserves further study, especially in domains where learning instances may coherently be grouped in multiple ways. Such efforts may have both interesting theoretical import as well as implications for the design of learning technology in applied settings.

Acknowledgments

We thank Tim Burke, Everett Mettler, Rachel Older, and research assistants from the UCLA Human Perception Lab. This material is based on work supported by the National Science Foundation under Grant No. DRL-1644916. Any opinions, findings, and conclusions are those of the authors and do not necessarily represent the views of the NSF.

References

Best, L.A., Smith, L.D. & Stubbs, D.A. (2007). Perception of linear and nonlinear trends: Using slope and curvature information to make trend discriminations. *Perceptual and Motor Skills, 104*, 707-721.

Bobko, P., & Karren, R. (1979). The perception of pearson product moment correlations from bivariate scatter plots. *Personnel Psychology, 32*(2), 313-325.

Boynton, D.M. (2000). The psychophysics of informal covariation assessment: Perceiving relatedness against a background of dispersion. *J. of Experimental Psychology: Human Perception and Performance, 26*(3), 867-876.

Doherty, M.E., & Anderson, R.B. (2009). Variation in scatter plot displays. *Beh. Res. Methods, 41*(1), 55-60.

Gibson, E.J. (1969). Principles of perceptual learning and development. New York: Appleton-Century-Crofts.

Kellman, P.J. (2002). Perceptual learning. In R. Gallistel (Ed.), *Stevens' handbook of exp. psych.: Learning, motivation & emotion*. (3rd ed.). NY: John Wiley & Sons.

Kellman, P.J., & Massey, C.M. (2013). Perceptual learning, cognition, and expertise. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 58). Amsterdam: Elsevier Inc.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science, 2*(2), 285-305.

Lauer, T. W. & Post, G. V. (1989). Density in scatter plots and the estimation of correlation. *Behaviour & Information Technology, 8*, 235 – 244.

Mettler, E. & Kellman, P.J. (2014). Adaptive response-time-based category sequencing in perceptual learning. *Vision Research, 99*, 111-123.

Mettler, E., Massey, C. M., & Kellman, P.J. (2016). A comparison adaptive and fixed schedules of practice. *J. of Experimental Psychology: General, 145*(7), 897 – 917.

Meyer, J., & Shinar, D. (1992). Estimation correlations from scatter-plots. *Human Factors*, *34*, 335 – 349.

Meyer, J., Taieb, M., & Flascher, I. (1997). Correlation estimates as perceptual judgments. *Journal of Experimental Psychology: Applied, 3*, 3 – 20.

Strahan, R. F., & Hansen, C. J. (1978). Underestimating correlation from scatter plots. *Applied Psychological Measurement, 2*, 543 – 550.

Thai, K. P., Krasne, S., & Kellman, P. J. (2015). Adaptive perceptual learning in electrocardiography: The synergy of passive and active classification. *Proc. of the 37th Annual Conference of the Cog. Sci. Society*. Austin, TX: Cognitive Science Society, 2350 – 2355.