Report

Current Biology

Perceptual Learning Generalization from Sequential Perceptual Training as a Change in Learning Rate

Highlights

- Training on multiple perceptual tasks produced significant learning generalization
- Learning generalization manifested as increased learning rate ("learning to learn")
- Standard methodology would tend to miss or misidentify this type of generalization

Authors

Florian Kattner, Aaron Cochrane, Christopher R. Cox, Thomas E. Gorman, C. Shawn Green

Correspondence

cshawn.green@wisc.edu

In Brief

The extent to which learning generalizes to new tasks is a key concern in the study of perceptual learning. Kattner, Cochrane, et al. report three experiments demonstrating that training on a series of tasks can induce generalization that manifests only in terms of increases in learning rate ("learning to learn"), not immediate performance.

Kattner et al., 2017, Current Biology *27*, 840–846
March 20, 2017
http://dx.doi.org/10.1016/j.cub.2017.01.046

CellPress

CellPress

Current Biology
Report

Perceptual Learning Generalization from Sequential Perceptual Training as a Change in Learning Rate

Florian Kattner,[1,3] Aaron Cochrane,[2,3] Christopher R. Cox,[2] Thomas E. Gorman,[2] and C. Shawn Green[2,4,*]
[1]Institute of Psychology, Technische Universität Darmstadt, Alexanderstr. 10, 64283 Darmstadt, Germany
[2]Department of Psychology, University of Wisconsin–Madison, 1202 West Johnson Street, Madison, WI 53706-1611, USA
[3]Co-first author
[4]Lead Contact
*Correspondence: cshawn.green@wisc.edu
http://dx.doi.org/10.1016/j.cub.2017.01.046

SUMMARY

With practice, humans tend to improve their performance on most tasks. But do such improvements then generalize to new tasks? Although early work documented primarily task-specific learning outcomes in the domain of perceptual learning [1–3], an emerging body of research has shown that significant learning generalization is possible under some training conditions [4–9]. Interestingly, however, research in this vein has focused nearly exclusively on just one possible manifestation of learning generalization, wherein training on one task produces an immediate boost to performance on the new task. For instance, it is this form of generalization that is most frequently referred to when discussing learning "transfer" [10, 11]. Essentially no work in this domain has focused on a second possible manifestation of generalization, wherein the knowledge or skills acquired via training, despite not being directly applicable to the new task, nonetheless allow the new task to be learned more efficiently [12–15]. Here, in both the visual category learning and visual perceptual learning domains, we demonstrate that sequentially training participants on tasks that share a common high-level task structure can produce faster learning of new tasks, even in cases where there is no immediate benefit to performance on the new tasks. We further show that methods commonly employed in the field may fail to detect or else conflate generalization that manifests as increased learning rate with generalization that manifests as immediate boosts to performance. These results thus lay the foundation for the various routes to learning generalization to be more thoroughly explored

RESULTS

Experiment 1: Generalization as a Change in Learning Rate in Novel-Shape Categorization

We have previously identified multiple distinct mechanisms that could, in principle, promote increases in the rate at which new tasks are learned (which has been referred to as the "learning to learn" form of generalization) without engendering any immediate benefit to performance [15–17]. In the present paper, we chose to focus on one of these mechanisms, which we have referred to as the "knowledge-based" mechanism. Here, through exposure to many different tasks that all share some higher-level structure or components, participants could potentially learn those regularities that exist across the individual tasks [14, 18–20]. Importantly, knowledge of such higher-level regularities need not provide any direct insight regarding how one should interpret or respond to stimuli when beginning a new task. Instead, this knowledge may serve to constrain or order the to-be-learned task space. If this is the case, such training will produce faster learning of new tasks without any immediate benefit to performance.

As an initial demonstration of the conditions that promote this form of generalization, we chose a domain where it is reasonably straightforward to induce the necessary higher-level shared task structure. While there are a number of domains where this would be possible, we selected novel-shape categorization. This choice was based primarily on the robust body of work outlining clear similarities between the novel-shape categorization domain and the perceptual learning domain, suggesting that the former could be a good model for the latter [21–26].

We first created a continuous 2D space from which individual novel shapes could be drawn. We then defined eight unique categories within that space, pairs of which could be utilized in four separate categorization learning tasks (Figure 1A; Supplemental Experimental Procedures). Importantly, while the individual shapes, categories, and category boundaries were unique to each learning task, many other higher-order features were shared across learning tasks. For instance, although the categories were placed in different parts of the 2D space in the different learning tasks, the general shape of the categories was shared (i.e., 2D Gaussians with similar parameters). Other shared aspects included the fact that the two to-be-discriminated categories in each learning task were always linearly separable and, furthermore, were always separable along a single dimension. Critically, although this shared structure provided no information that would be immediately applicable in a new task, it should nonetheless allow new tasks that share this same structure to be learned more quickly.

Twenty-four participants underwent this series of four novel-shape categorization learning tasks (60 trials each). In examining their behavior, we first assessed whether any immediate

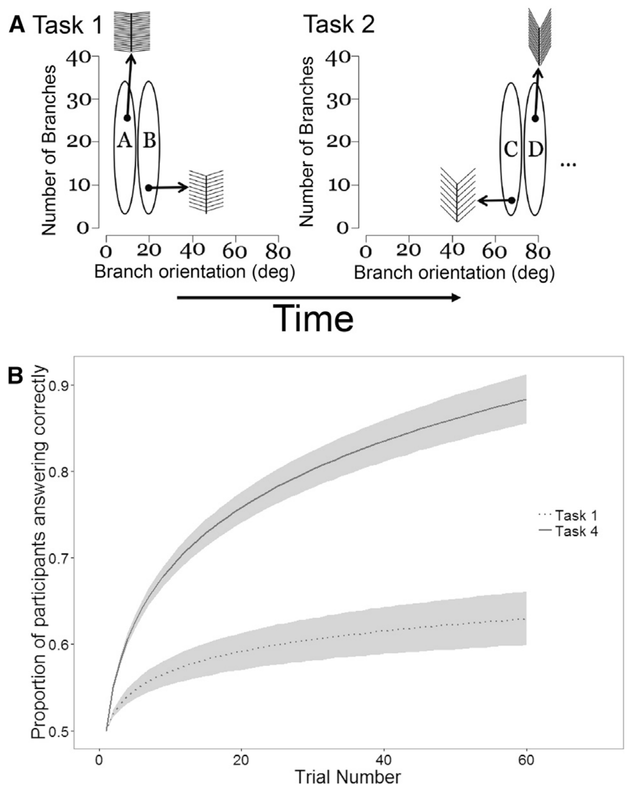

Figure 1. Sequential Novel Shape Categorization Task and Results

(A) Novel shapes ("feathers") were drawn from a 2D space, where one dimension corresponded to the number of branches on the feather and the other dimension corresponded to the orientation of the branches (e.g., a point in the top left of this space produces a feather with many branches with a very flat orientation, while a point in the bottom right of this space produces a feather with few branches with a very steep orientation). For each learning task, feathers were drawn from one of two categories defined by 2D Gaussians in the space (e.g., category A versus category B in the left panel and category C versus category D in the right panel). While the various tasks involve totally different shapes, categories, and category boundaries, they share a certain degree of high-level structure, including the general shape of the categories in the space and the fact that the discriminant that best separates the categories lies along a single dimension.

(B) Although participants started with the same (chance-level) performance in both their first and final categorization tasks, they learned much more quickly in the final categorization task as compared to the first (note that, for correspondence with experiment 2, we plotted only the first and final categorization tasks here; see Figure S1 for behavioral performance across all four categorization tasks and Figure S2 for the bootstrapped learning slope estimates across all four categorization tasks). Error bands represent 95% confidence intervals.

changes in performance were observed from task to task by examining first-trial performance across the four tasks. Consistent with our expectation that no such immediate changes would be observed, participants began with similar, chance-level performance in all four tasks (task 1: M = 0.42; task 2: M = 0.54; task 3: M = 0.54; task 4: M = 0.46; none of these values were significantly different either from one another or from chance; Figures 1B and S1). We next tested the hypothesis that participants would learn more quickly as they progressed from task to task. To this end, we computed bootstrapped estimates of the learning curve for each task (see the Supplemental Experimental Procedures) and then assessed the extent to which the rate of learning was increasing from task to task. Consistent with our hypothesis, we found that participants did indeed learn to categorize more quickly as they moved from task to task (e.g., the slope of the average best fitting line to the learning rates across tasks was significantly above zero, indicating faster learning from task to task: $b = 0.026$, $p < 0.001$; see Figures 1B, S1, and S2 and the Supplemental Experimental Procedures for additional quantification). These results thus strongly indicate that properly sequenced training on tasks containing shared higher-order structure can induce generalization in the form of learning rate, in the absence of any immediate changes in performance.

Experiment 2: Generalization as a Change in Learning Rate in Perceptual Learning

Given the results described above, we next applied similar logic to the domain of perceptual learning. Specifically, we first designed a set of five perceptual learning tasks—visual grating spatial frequency categorization, color lightness categorization, dot bisection, Gabor orientation categorization, and dot motion direction categorization. These tasks shared little in terms of the base features of the stimuli utilized in the various tasks (e.g., the stimuli included visual gratings of various spatial frequencies, square color patches of various lightness, three roughly vertically arranged dots, visual Gabors of various orientations, and fields of dots moving in various average directions; some stimuli were presented centrally, and some stimuli were presented peripherally at different spatial positions depending on the task; see the Supplemental Experimental Procedures). Note that the particular base features were chosen partially because they are dimensions along which learning specificity has been commonly observed in the perceptual learning literature (i.e., there are many examples of learning that failed to generalize across position, spatial frequency, orientation, motion direction, etc.; for a review, see [27]).

Despite the lack of similarity at the level of exact stimuli, the tasks in experiment 2 nonetheless shared a great deal of higher-level structure. For instance, the base timing structure, including a 150-ms stimulus presentation followed by a 500-ms mask, was shared across all tasks. Other higher-level aspects that were shared across tasks included the fact that stimuli were always drawn from a uniform distribution and that the category boundary was always found in the center of that uniform distribution. As was true in experiment 1, the shared structure across tasks thus provided no information regarding the exact choice that should be made on the first trial of a new task (e.g., none of the shared structure indicated whether an observed Gabor was clockwise or counterclockwise relative to a given reference angle). However, the shared structure could, for instance, provide information that would allow the participants to more quickly learn to separate signal from noise (and thus improve performance more quickly overall).

Thirteen participants were trained sequentially on the five different perceptual learning tasks (first four tasks: 800 trials per day, 2 days each; fifth task: 500 trials on a single day). Meanwhile, a second cohort of ten participants underwent only the final task (note that we refer to the comparison between these

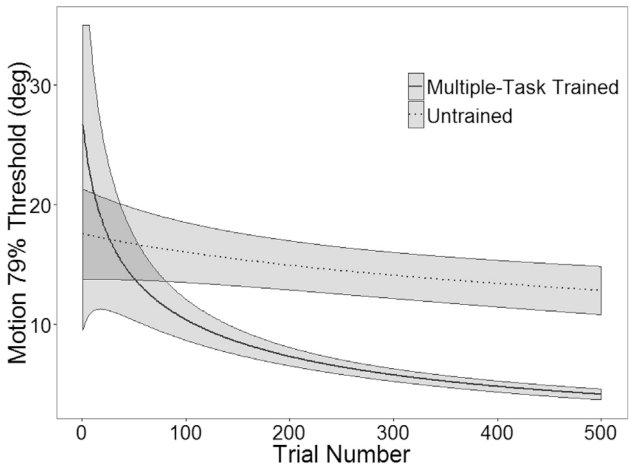

Figure 2. "Learning to Learn" without "Transfer" in Perceptual Learning

Although both the multiple-task trained and untrained participants started with identical initial levels of performance on the dot motion direction categorization task, the trained individuals (i.e., those participants who had previously undergone perceptual learning on four tasks with similar high-level structure) learned much faster (see Figure S4 for fitting method comparison and individual-level data, Figure S3 for the first/fourth comparison, and Table S1 for details on the second and third trained tasks). Error bands represent 95% confidence intervals.

two groups as the "multiple-task trained/untrained comparison"). This setup allowed us to then directly assess whether the multiple-task trained participants showed differences in either initial performance or learning rate on the final perceptual learning task as compared to the untrained participants. Furthermore, in order to allow for additional, within-participant tests of our hypotheses, six of the multiple-task trained participants performed the first four perceptual learning tasks in one order, while the remaining seven performed the same tasks in the reverse order (see the Supplemental Experimental Procedures). By combining performance (via *Z* scoring) across the participants' respective first training tasks (six participants: spatial frequency; seven participants: orientation) and the participants' respective fourth training tasks (vice versa), we could make a similar set of comparisons as in the multiple-task trained/untrained case but within participants rather than between participants (we refer to this as the "first/fourth comparison" below).

For each participant and task, the data were fit via a time-evolving logistic function that has previously been used by our group to examine learning in the perceptual domain [7, 8]. Critically, unlike standard fitting techniques that aggregate over large blocks of trials, our statistical approach allows for an estimate both of immediate changes in performance (making use of the estimated threshold on the first trial of each new task) and of changes in learning rate (making use of the rate at which the psychometric function changes over time). Our prediction was that for both the multiple-task trained/untrained and first/fourth comparisons, we would see no differences in first-trial performance but that we would see significant differences in learning rate.

As can be seen in Figure 2 (see also Figure S3), both hypotheses were confirmed. No significant difference in first-trial performance was seen in either the multiple-task trained/untrained comparison (untrained: threshold = 17.55 ± 1.88; multiple-task trained: threshold = 26.71 ± 8.61; $t(13.1) = 1.04$, $p = 0.32$) or the first task/fourth task comparison (first task: *Z*-scored threshold = 1.36 ± 0.38; fourth task: *Z*-scored threshold = 0.85 ± 0.25; $t(12) = 0.99$, $p = 0.34$). There was, however, a clear difference in the rate at which participants learned in both the multiple-task trained/untrained comparison (rate of change for the psychometric function: for untrained, $5.29 \times 10^{-5} \pm 1.60 \times 10^{-5}$; for multiple-task trained, $4.85 \times 10^{-4} \pm 6.06 \times 10^{-5}$; $t(13.7) = 6.9$; $p < 0.001$; see the Supplemental Experimental Procedures) and the first/fourth comparison (*z*-scored rate of change for the psychometric functions: for first tasks, −0.59 ± 0.13; for fourth tasks, 0.59 ± 0.24; $t(12) = 5.7$, $p < 0.001$). Therefore, as was true in experiment 1, the data clearly indicate that sequential training on multiple perceptual learning tasks that share higher-order structure can induce changes in learning rate in the absence of any immediate changes in performance.

Experiment 3: Assessing the Role of Training Variety

Two key questions are raised by the results of experiment 2. The first question is whether the results were dependent upon a training regimen that included multiple tasks or whether the same effect would be induced by training on a single task for the same total amount of time. Many theoretical frameworks [15, 28], suggest a critical role for variety of experience, as encountering the same higher-level structure in multiple different tasks/contexts is a cue that the structure is indeed more broadly applicable. However, it is certainly possible that experience with the same statistical structure in a single exemplar task would be equivalently valuable.

The second question is whether the observed enhancement in the learning of new tasks is indeed dependent upon shared statistical structure. The results of experiment 2 clearly show that after learning a number of tasks that share the same structure, participants are able to learn a new task that shares that same learned structure more quickly. However, an additional explicit prediction from our framework is that if this learned structure is violated in a new task, then performance should suffer (i.e., would be worse than if participants hadn't completed any previous learning tasks).

To address these issues, we again trained nine new participants sequentially on five different perceptual learning tasks. Unlike in experiment 2, though, these participants began by completing a total of 6,400 trials of the initial orientation training task (800 trials per day, 8 days; i.e., the same number of total trials/days as for the first four training tasks in experiment 2; as such, this group is referred to as a the "single-task trained group"; see the Supplemental Experimental Procedures). The participants then completed 800 trials of the same motion task as in experiment 2. Comparing the first 500 trials of motion task performance of the participants in experiment 3 with those in experiment 2 thus offers a clear assessment of the role of variety in the "learning to learn" effect (i.e., both groups would have completed 6,400 trials of perceptual learning prior to completing the motion learning task— experiment 2 participants having done so across four tasks and experiment 3 participants having done so across a single task). After the motion task, participants completed the same basic color lightness categorization and dot bisection tasks from experiment 2. This was done in order to

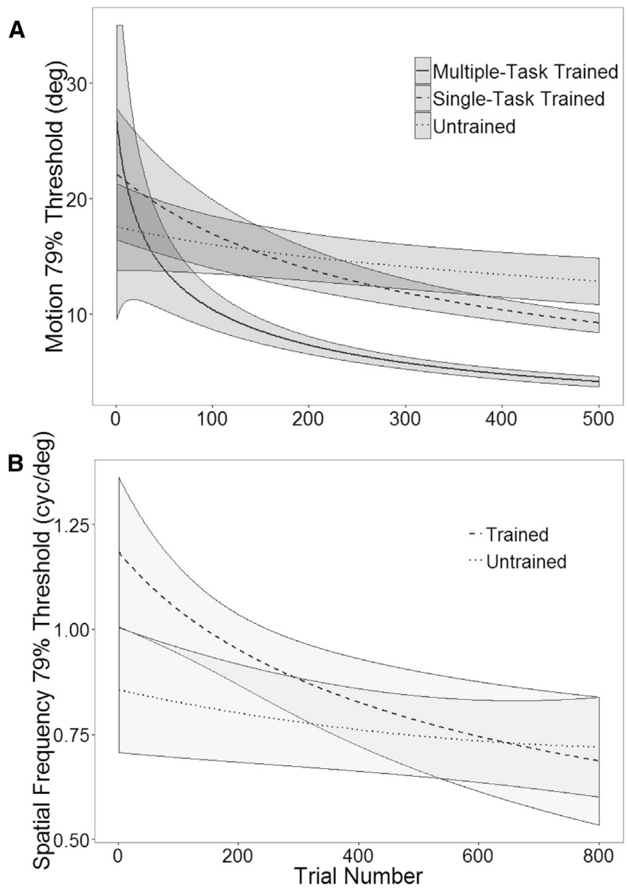

Figure 3. Assessing the Role of Training Variety and Violations of Task Structure

(A) The performance of the single-task trained group in experiment 3 on the motion learning task was intermediate to both of the groups from experiment 2. Learning was significantly faster than that of the untrained group but slower than that of the multiple-task trained group. Error bands represent 95% confidence intervals.

(B) The violations of the learned-task structure in the spatial frequency learning task resulted in poorer initial performance in the trained group of experiment 3 when compared to the untrained group. However, there was a trend for the learning rate to still be faster in the trained group than in the untrained group. This would be consistent with the fact that, while one aspect of the training was violated in the spatial frequency task (i.e., the exact temporal order/structure), many other aspects remained shared (e.g., the fact that some aspects of the presentation were stimuli, whereas some were noise; the fact that the stimulus was quickly presented; the fact that the stimuli differed along a single continuous dimension with the category boundary lying in the center of the uniform distribution; etc.). Error bands represent 95% confidence intervals.

ensure that the trained participants in experiment 3 experienced the higher-level task structure with the same amount of variety as the participants in experiment 2 before completing a final task aimed at addressing the second key question above (i.e., whether the enhanced learning was dependent upon new tasks that share the same structure as learned tasks). To this end, as a final task, participants completed a new version of the spatial frequency task that was redesigned from the experiment 2 version to violate one major aspect of the previous training. Specifically, in this new version, the random 500-ms mask came before the stimulus, rather than after. If one key piece of information taught by the preceding training tasks was the temporal relationship between stimulus (target) and noise (mask), this final task would be expected to be more difficult for the participants who had undergone the training than for another new group of participants (n = 9), who only performed the spatial frequency task.

Answering the question of the role of variety, the learning rate parameters in the motion learning task were significantly different between the single-task trained group from experiment 3 and both the untrained group from experiment 2 and the multiple-task trained group from experiment 2 (Figure 3A). Specifically, the single-task trained group learned significantly more quickly than the untrained group (for the single-task trained group, $1.59 \times 10^{-4} \pm 1.62 \times 10^{-5}$; for the untrained group, $5.29 \times 10^{-5} \pm 1.60 \times 10^{-5}$; $t(16.9) = 4.66$; $p < 0.001$) but significantly more slowly than the multiple-task trained group (for the multiple-task trained group, $4.85 \times 10^{-4} \pm 6.06 \times 10^{-5}$; $t(13.7) = 5.2$, $p < 0.001$).

For the question of the effect of violated task structure, as expected, repeated training on tasks with similar statistics led to a decreased initial performance when these statistics were violated (Figure 3B). Single-task trained participants had significantly higher initial spatial frequency thresholds than participants who had not completed any training (we use the label "single-task trained" in order to differentiate the trained group in experiment 3 from the trained group in experiment 2; however, note again that the single-task trained group had in fact been trained on four separate tasks before they were trained on the spatial frequency task: for single-task trained, 1.18 ± 0 0.089; for untrained, 0.86 ± 0 0.074; $t(16.1) = 2.8$; $p = 0.012$). As shown in Figure 3B, this difference in initial threshold was accompanied by a non-significantly faster learning rate for single-task trained versus untrained participants (for single-task trained, $1.21 \times 10^{-3} \pm 4.03 \times 10^{-4}$; for untrained, $3.74 \times 10^{-4} \pm 2.46 \times 10^{-4}$; $t(13.4) = 1.8$, $p = 0.098$).

While the results of both questions are broadly consistent with the expectations of the our theoretical framework, caution is warranted when attempting to interpret the exact changes that gave rise to the final learning task performance. For instance, one interpretation of the motion learning task data in experiment 3 involves only changes in a positive direction—that is, that the single-task training led to some learning of the shared structure, but it was of a lesser degree than the learning induced by the multiple-task training. An alternative interpretation, though, would involve changes in both a positive direction (some learning of the shared structure) and a negative direction (i.e., stronger learning of the non-shared task structure, such as the spatial locations that attention should be guided toward). Similarly, while initial performance on the spatial frequency learning task in experiment 3 was worse for trained than for untrained participants (consistent with violated expectations), there was nonetheless a strong trend toward faster learning (that would be consistent with the fact that, although one task aspect was violated in the spatial frequency task—i.e., the exact timing—many other task aspects were still shared—e.g., the fact that critical information was briefly presented, that some of the presentation was pure noise, etc.). This is particularly critical, as it is unclear that "unlearning" previously learned information should follow the same temporal dynamics (but in an opposite direction) as the initial learning [29].

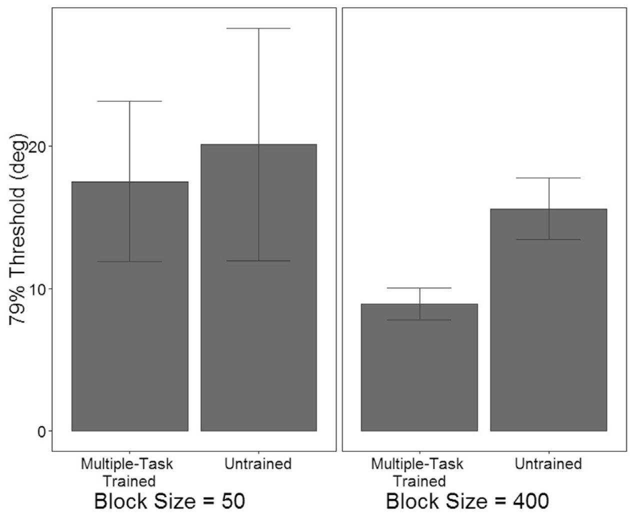

Figure 4. Inferences about Generalization that Would Have Been Drawn via More Standard Techniques

Although the data analysis technique we employed in experiments 2 and 3 specifically models changes in performance as a continuous function of time, it is considerably more standard in the literature to fit data as a single block. However, if this block is too small (left bars), no difference between groups may be detected (i.e., differences that would have been present due to "learning to learn" would be missed). Conversely, if this block is too large (right bars), generalization will be detected, but it will be wrongly identified as "transfer" rather than "learning to learn" (i.e., without modeling performance as a function of time, there is no way to determine whether the observed differences were present immediately or evolved through time). See also Figure S4 for a comparison of fitting methods. Error bars represent 95% confidence intervals.

DISCUSSION

The results of the present investigation clearly demonstrate that properly designed sequential training can induce perceptual learning generalization that manifests as a change in learning rate, in the absence of any immediate changes in performance. This is consistent with the broad idea that learning higher-level structure can, in turn, facilitate learning the individual parameters of new tasks, thus inducing what has been called, in various parts of the literature, "learning to learn" [14–17, 20, 28, 30–33]. Critically, this is not a simple matter of directly applying some known information to a new task (either immediately or delayed through time), which is commonly referred to as "transfer" of learning. Indeed, in our case here, the higher-level structure that exists across tasks (e.g., the consistent timing information) provides no information that would directly inform actions in each new task (i.e., the timing information provides no information regarding what separates "high" from "low" spatial frequency responses).

In order to explore this distinction further, significant methodological changes may be necessary for the perceptual learning field going forward. For instance, one of the more common designs used to examine perceptual learning generalization involves training on some perceptual learning task "A" followed by a single block of some generalization task "B." Unfortunately, this type of design may result in the "learning to learn" form of generalization being missed entirely, or else it may result in the "learning to learn" form of generalization being mislabeled as immediate transfer (depending on the number of trials used and the rate at which the task is learned). To make this issue explicit, we assessed the inferences that would have arisen had we utilized more typical methodological and statistical approaches in experiment 2. In the first case, we fit the data for the first 50 trials of the generalization tasks (mimicking a very short generalization task), while in the second case, we fit the data over the first 400 trials of the generalization tasks (in both cases, the data were also fit in a more conventional fashion—i.e., aggregating over the entire block of trials rather than explicitly modeling performance changes as a function of time—to demonstrate that any outcomes were not specific to the analysis technique).

As can be seen in Figures 4 and S4, when examining just the first 50 trials of data for the generalization task, no significant differences were observed (threshold over first 50 trials: multiple-task trained/untrained comparison—for untrained, 20.12 ± 4.07; for multiple-task trained, 17.52 ± 2.80; t(16.71) = 0.52, p = 0.61; note that a similar outcome is seen in the first/fourth comparison: for first tasks, −0.14 ± 0.20; for fourth tasks, 0.14 ± 0.33; $t(12) = 0.70$, $p = 0.49$). In other words, this approach would have led to the correct conclusion that no "transfer" was present, but due to the small number of trials involved, it would have resulted in a failure to detect the presence of "learning to learn" (i.e., there would not have been sufficient time for the groups to split apart). Conversely, when examining the first 400 trials as a single block, significant differences were observed for both the trained/untrained and first/fourth comparisons (threshold over first 400 trials: trained/untrained comparison—for untrained, 15.59 ± 1.08; for multiple-task trained, 8.92 ± 0.57; $t(13.9) = 5.47$, $p < 0.001$; first/fourth comparison—for first tasks, 0.42 ± 0.29; for fourth tasks, −0.42 ± 0.19; $t(12) = 2.90$, $p = 0.01$). In essence, by aggregating performance over a large number of trials, one would correctly infer that learning generalization was present but would erroneously conclude that this reflects "transfer" rather than differences in learning rate (i.e., without modeling performance across time, there is no way to differentiate an immediate difference in performance from a rapid splitting apart of performance). This latter issue might be compounded by the use of staircase procedures to estimate thresholds (a common procedure in the field), because the single data point that arises from a staircase procedure is the result of tens, if not hundreds, of trials. Finally, while it is common to utilize some number of practice trials prior to the actual generalization task, this too may result in issues with correctly identifying the form of generalization that is present, as the practice trials could provide an opportunity for two groups to begin splitting apart (in which case, group differences may then be observed as early as the first few trials of the actual generalization task).

We note, though, that although this framework makes a clear prediction that learning multiple tasks with shared task structure should increase learning rate on new tasks that share the same structure, the predictions regarding first-trial performance are much more task dependent. The training tasks in the experiments above were designed in order to minimize the extent to which the shared structure should inform first-trial performance, but this need not be the case. In situations where, for instance, participants bring some knowledge about the new task dimensions, one might expect to see both better initial performance and faster learning (i.e., if there is a match between both the higher-level structure and some number of the task-level parameters). The types of methodological changes suggested above could thus pave the way toward further addressing what is always the key question for the field—namely, ''what'' is being learned from training on a given task. The evidence provided here strongly indicates that considering performance on new untrained tasks, both in terms of immediate performance and in terms of learning rate, can serve to differentiate and identify what has been learned via training in a way not provided for by previous methods (e.g., in both ''transfer'' and ''learning to learn,'' one is applying previous experience to new tasks, but in the former case, that knowledge is directly applicable to the new task and thus benefits appear immediately, while in the latter case, the knowledge can only serve to shape learning of the new task).

The current data may also suggest the need to further explore a number of other previous results, both within and beyond the domain of perceptual training. For instance, there is a great deal of research on the impact that various complex forms of experience have on perceptual and cognitive skills (e.g., cognitive training, ''brain training,'' etc. [34–37]). As has generally been true of the standard perceptual learning literature, work in these fields has focused exclusively on the ''transfer'' form of generalization without necessarily utilizing methods that could differentiate ''transfer'' from ''learning to learn.'' It is thus possible that ''learning to learn'' has, at times, been misidentified as ''transfer'' (in the case of positive results) and/or that the ''learning to learn'' form of generalization was not detected (in the case of non-significant generalization results). The same is also potentially true of work on more complex training regimens within the perceptual learning domain, such as the nicely elaborated ''rules-based learning'' framework [6]. Here, depending on the ''rule'' that is learned, one might expect *either* immediate ''transfer''—as, for instance, would be predicted if the rule were essentially a template for the to-be-identified target—or ''learning to learn''—as would be predicted if the rule was, for instance, more broadly about how to best separate targets from noise. And if rules at various levels of abstraction are learned [38–40], this could result in *both* some degree of immediate ''transfer'' and some degree of ''learning to learn'' (or, indeed, one could imagine a situation where one must learn some task statistics before knowing which of several possible rules to adopt). Examining these issues further, though, will require the types of methodological designs and statistical analyses highlighted here that can separate these distinct forms of generalization.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2017.01.046.

AUTHOR CONTRIBUTIONS

Conceptualization, C.S.G.; Methodology, C.S.G., F.K., A.C., and T.E.G.; Investigation, C.S.G., F.K., and T.E.G.; Analysis, A.C. and F.K.; Writing – Original Draft, C.S.G., F.K., A.C., and C.C.; Writing – Review & Editing, C.S.G., F.K., A.C., C.C., and T.E.G.; Funding Acquisition, C.S.G.; Supervision, C.S.G. and F.K.

ACKNOWLEDGMENTS

This work was supported by Office of Naval Research grants N00014-14-1-0512 and N00014-17-1-2049 to C.S.G. and a University of Wisconsin Office of the Vice Chancellor for Research and Graduate Education Research Fall Research Competition award to C.S.G. The research involved the participation of human subjects and it was approved by the University of Wisconsin-Madison Education and Social/Behavioral Sciences Institutional Review Board. All participants provided written informed consent prior to participation.

Received: July 27, 2016
Revised: December 5, 2016
Accepted: January 23, 2017
Published: March 2, 2017

REFERENCES

1. Fiorentini, A., and Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. Nature *287*, 43–44.
2. Ball, K., and Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. Science *218*, 697–698.
3. Fahle, M. (2004). Perceptual learning: a case for early selection. J. Vis. *4*, 879–890.
4. Jeter, P.E., Dosher, B.A., Petrov, A., and Lu, Z.L. (2009). Task precision at transfer determines specificity of perceptual learning. J. Vis. *9*, 1–13.
5. Xiao, L.Q., Zhang, J.Y., Wang, R., Klein, S.A., Levi, D.M., and Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. Curr. Biol. *18*, 1922–1926.
6. Zhang, J.Y., Zhang, G.L., Xiao, L.Q., Klein, S.A., Levi, D.M., and Yu, C. (2010). Rule-based learning explains visual perceptual learning and its specificity and transfer. J. Neurosci. *30*, 12323–12328.
7. Snell, N., Kattner, F., Rokers, B., and Green, C.S. (2015). Orientation transfer in vernier and stereoacuity training. PLoS ONE *10*, e0145770.
8. Green, C.S., Kattner, F., Siegel, M.H., Kersten, D., and Schrater, P.R. (2015). Differences in perceptual learning transfer as a function of training task. J. Vis. *15*, 5.
9. Deveau, J., Ozer, D.J., and Seitz, A.R. (2014). Improved vision and on-field performance in baseball through perceptual learning. Curr. Biol. *24*, R146–R147.
10. Thorndike, E.L., and Woodworth, R.S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. Psychol. Rev. *8*, 247–261.
11. Singley, M.K., and Anderson, J.R. (1989). The Transfer of Cognitive Skill (Harvard University Press).
12. Harlow, H.F. (1949). The formation of learning sets. Psychol. Rev. *56*, 51–65.
13. Thrun, S., and Pratt, L.e. (1998). Learning to Learn (Kluwer Academic Publishers).
14. Kemp, C., Goodman, N.D., and Tenenbaum, J.B. (2010). Learning to learn causal models. Cogn. Sci. *34*, 1185–1243.
15. Bavelier, D., Green, C.S., Pouget, A., and Schrater, P. (2012). Brain plasticity through the life span: learning to learn and action video games. Annu. Rev. Neurosci. *35*, 391–416.
16. Braun, D.A., Mehring, C., and Wolpert, D.M. (2010). Structure learning in action. Behav. Brain Res. *206*, 157–165.
17. Bejjanki, V.R., Zhang, R., Li, R., Pouget, A., Green, C.S., Lu, Z.L., and Bavelier, D. (2014). Action video game play facilitates the development of better perceptual templates. Proc. Natl. Acad. Sci. USA *111*, 16961–16966.
18. Botvinick, M.M. (2008). Hierarchical models of behavior and prefrontal function. Trends Cogn. Sci. *12*, 201–208.
19. Hinton, G.E. (2007). Learning multiple layers of representation. Trends Cogn. Sci. *11*, 428–434.

20. Michel, M.M., and Jacobs, R.A. (2007). Parameter learning but not structure learning: a Bayesian network model of early perceptual learning. J. Vis. *7*, 4.
21. Edelman, S., and Duvdevani-Bar, S. (1997). A model of visual recognition and categorization. Philos. Trans. R. Soc. Lond. B Biol. Sci. *352*, 1191–1202.
22. McLaren, I.P.L. (1997). Categorization and perceptual learning: an analogue of the face inversion effect. Q. J. Exp. Psychol. A *50*, 257–273.
23. Mitchell, C., and Hall, G. (2014). Can theories of animal discrimination explain perceptual learning in humans? Psychol. Bull. *140*, 283–307.
24. Garrigan, P., and Kellman, P.J. (2008). Perceptual learning depends on perceptual constancy. Proc. Natl. Acad. Sci. USA *105*, 2248–2253.
25. Kellman, P.J. (2004). Perceptual learning. In Stevens' Handbook of Experimental Psychology, Third Edition, H. Pashler, ed. (Wiley).
26. Kattner, F., Cox, C.R., and Green, C.S. (2016). Transfer in rule-based category learning depends on the training task. PLoS ONE *11*, e0165260.
27. Sagi, D. (2011). Perceptual learning in vision research. Vision Res. *51*, 1552–1566.
28. Braun, D.A., Aertsen, A., Wolpert, D.M., and Mehring, C. (2009). Motor task variation induces structural learning. Curr. Biol. *19*, 352–357.
29. Dickinson, A., and Pearce, J.M. (1977). Inhibitory interactions between appetitive and aversive stimuli. Psychol. Bull. *84*, 690–711.
30. Griffiths, T.L., and Tenenbaum, J.B. (2005). Structure and strength in causal induction. Cognit. Psychol. *51*, 334–384.
31. Brown, A.L., and Kane, M.J. (1988). Preschool children can learn to transfer: learning to learn and learning from example. Cognit. Psychol. *20*, 493–523.
32. Tenenbaum, J.B., and Griffiths, T.L. (2001). Generalization, similarity, and Bayesian inference. Behav. Brain Sci. *24*, 629–640.
33. Green, C.S., and Bavelier, D. (2012). Learning, attentional control, and action video games. Curr. Biol. *22*, R197–R206.
34. Jaeggi, S.M., Buschkuehl, M., Jonides, J., and Perrig, W.J. (2008). Improving fluid intelligence with training on working memory. Proc. Natl. Acad. Sci. USA *105*, 6829–6833.
35. Jaeggi, S.M., Buschkuehl, M., Jonides, J., and Shah, P. (2011). Short- and long-term benefits of cognitive training. Proc. Natl. Acad. Sci. USA *108*, 10081–10086.
36. Kellman, P.J., Massey, C.M., and Son, J.Y. (2010). Perceptual learning modules in mathematics: enhancing students' pattern recognition, structure extraction, and fluency. Top. Cogn. Sci. *2*, 285–305.
37. Anguera, J.A., Boccanfuso, J., Rintoul, J.L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., and Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. Nature *501*, 97–101.
38. Tenenbaum, J.B., Kemp, C., Griffiths, T.L., and Goodman, N.D. (2011). How to grow a mind: statistics, structure, and abstraction. Science *331*, 1279–1285.
39. Kemp, C., Perfors, A., and Tenenbaum, J.B. (2007). Learning overhypotheses with hierarchical Bayesian models. Dev. Sci. *10*, 307–321.
40. Lake, B.M., Salakhutdinov, R., and Tenenbaum, J.B. (2015). Human-level concept learning through probabilistic program induction. Science *350*, 1332–1338.