International Journal of Engineering Education Vol. 36, No. 3, pp. 1082–1096, 2020
Printed in Great Britain

0949-149X/91 $3.00+0.00

Visual Expertise in Fluid Flows: Uncovering a Link Between Conceptual and Perceptual Expertise*

KATHERINE GOODMAN
Inworks, University of Colorado Denver, Denver, CO, USA. Corresponding author: Katherine Goodman, P.O. Box 173364, CB 119A, Denver, CO80217, USA. E-mail: katherine.goodman@ucdenver.edu

JEAN HERTZBERG
Department of Mechanical Engineering, University of Colorado Boulder, Boulder, CO, USA.

TIM CURRAN
Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA.

ERIN AUSTIN
Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO, USA.

In previous research on fluid mechanics courses, students expressed engagement by relating moments of noticing fluid phenomena in everyday life. This implies learning transfer and cognitive flexibility about fluids. This study lays the foundation for connecting perceptual experiences with conceptual understanding, with implications for sensory-rich learning experiences. Emulating cognitive psychology experiments in visual expertise, we tested two groups of participants: "novices" (no formal fluids education, n = 56), and "experts" (passed at least one fluid mechanics course, n = 36). Without being told the categories, participants were trained to sort static images of fluid flows as laminar or turbulent. Half the participants in each group trained on flow images with a specific format (Von Kármán vortex streets), half on a more varied group of flow images. Participants were then tested on the same type of images as their training (post-test) and tested again on images from the other training group (alt-test). Training resulted in statistically significant gains for all four participant groups, comparing post-test to pre-test. An ANOVA of between-group differences revealed that experts did significantly better than novices ($p = 0.0266$), whereas a comparison by training-type was not significant ($p = 0.2758$). A comparison of alt-test to pre-test data revealed that learning generalized for Novices trained on General Images ($p = 0.0266$) but not for Novices trained on Vortex Streets ($p = 1.0$). Expert results non-significantly trended toward learning generalization. Despite inconclusive results on expert learning, this study provides a new direction to explore the learning of fluids and other constraint-based interactions.

Keywords: Knowledge transfer; experimental research; mechanical engineering; expert-novice; conceptual learning

1. Introduction

When we consider the effectiveness of education, we often ask whether students can transfer knowledge and skills from the classroom to other settings. This question is sometimes called the "transfer problem" [1], but it also appears under other names including "generalization" [2], "awareness"[3], or the "need to activate resources" [4]. Instructors who are aware of the situated nature of learning can intentionally develop learning environments that provide appropriate scaffolding for students [5, 6], making explicit for students which elements are necessary for the work and which others are incidental. However, without this realization, the contextual backdrop can become a barrier, inhibiting students from applying their new knowledge or skills in other contexts, defined by Engeström [7] as encapsulation. As a result, engineering students who have learned primarily through structured problem sets may not know how to apply that knowledge once in the workplace [8], and misconceptions that they hold as students often persist into their years as professionals [9].

Fluid mechanics, in particular, is a field in which students must apply what they have learned in a wide range of situations, from common mechanical engineering problems (e.g., pipe flow) to aerospace ones (aerodynamics of airfoils) to interdisciplinary applications in areas such as biomedical engineering, requiring greater cognitive flexibility [10]. Many of the processes in fluids are what Chi calls *constraint-based interactions* (CBIs) [11], which are ontologically different from the more common *event* or *system* processes. Fluid phenomena, such as diffusion, to use Chi's example, may appear to have a beginning and an end, but in fact are continuous. The surface appearance of a system belies a more complex, ongoing, and deeper mechanism, which may be part of why many students struggle with the concepts, especially once these ongoing fluid CBIs are part of larger engineered systems. Chi describes gaining insight about these CBIs as a shift across ontological

* Accepted 15 January 2020.

categories, and this often creates an ''aha!'' moment, very similar to moments of deep creative insight, ''whereby everything all of a sudden seems to make sense'' [11, p. 230]. If we accept Chi's definition of creativity (''the ability to re-represent a concept that one has to understand from one perspective to a 'fundamentally different' perspective'' [11, p. 230]), then learning fluid mechanics often requires a level of creativity some disciplines do not.

One learning theory that grapples well with both this complexity and the transfer problem is Cognitive Flexibility Theory (CFT) [10], [12]. Particularly for ill-structured domains, CFT acknowledges the need to ''revisit the same material, at different times, in rearranged contexts, for different purposes, and from different conceptual perspectives'' [10, p. 65]. As noted above, fluid mechanics is just such a knowledge domain, requiring multiple real-world instances of its application for students to be able to apply it in their future work. CFT posits that ''revisiting [an idea] is not repeating'' it [12, p. 6]. This has the pleasing property of aligning with insights into the neurobiology of remembering and forgetting. Each time a person recalls something, there is a moment when the weights of the neurons are changed, when the memory is literally edited [13], a phenomena that has been noted in long term studies [14]. These findings set CFT as a biologically plausible learning theory.

Furthermore, CFT has been posited as a reason why *transformative experiences* create enduring learning [15]. *Transformative experience* is a concept that stems from Dewey's seminal theory of experiential learning and is influenced by his work on the value of aesthetic experiences [16–18]. Simply put, students' perception of the world should change as a result of new knowledge and abilities. Transformative experience ties together learning and motivation, and is usually detected through three hallmarks; (1) Students naturally relate course concepts to what they see in the larger world (*expanded perception*), (2) they put those concepts to work (*motivated use*), and (3) they find value or meaning in that experience (*experiential* or *affective value*). Numerous studies in science education have used surveys related to transformative experience as measures of student learning and engagement [18–21].

Expanded perception, as required by the transformative experience, may be that moment of ''flexible reassembly of preexisting knowledge to adaptively fit the needs of a new situation'' as described by CFT [10, p. 59]. In previous research on fluids courses in a mechanical engineering department [22], we found instances of expanded perception. Students often expressed engagement by relating their visual experiences with fluids: ''I'll never ignore the sky again'' and ''I see examples of flow vis all the time now.'' These unsolicited self-reports prompted further study on the course itself and also the current study into the relationship between perception and conceptual understanding in a particular engineering content area.

Our work is motivated by a desire to quantify the underlying ability represented by these self-reported expansions of perception. Our investigation is informed by studies of *perceptual expertise*, which include research in face recognition and experts' perception and memory of cars and birds [23, 24]. *Perceptual expertise* is the context-sensitive, trainable ability to actively notice (or perceive) something through our senses and includes the ability to differentiate very similar items into their correct categories and sub-categories. Our experiment is modeled on studies in which novices were trained to recognize and classify similar stimuli, such as species of wading birds and owls [25–28]. While work on how people form concepts based on experience has long been an emphasis in psychology [29], these more recent studies, from experimental psychology and cognitive neuroscience, shift the focus to understanding the neurobiological source of perception. To date, cognitive psychology work in visual expertise has focused on concrete stimuli rather than concepts. Our ultimate goal, in contrast, would be to connect students' reports of expanded perception in fluids to a measurable ability: perceptual expertise in fluid physics.

To begin, we aimed to define and measure perceptual expertise for a specific characteristic of fluids: laminar versus turbulent flow. This expertise is somewhat different from that of content with more discrete category boundaries, such as bird species or car make and model classifications, which have been used in prior visual expertise studies. This is because the laminar/turbulent distinction represents a physical concept that can apply to nearly every fluid flow. Note that the need to maintain either turbulent or laminar flow is important in many engineering designs, and therefore a crucial concept. For the purposes of this study, laminar and turbulent can be treated as distinct categories because we removed any images that might be considered transitional or uncertain. Details that signal an image is of turbulent flow include irregular features at varying spatial scales. For laminar flow, the details of note include smooth layers, no sign of cross currents, and features all being of similar scale. Both laminar and turbulent flow may have visually identifiable shapes in common, including large scale vortices. This can make classification more complex for participants to learn.

Real-world expertise takes months or years to develop, yet we can still gain preliminary insight on the mechanisms of learning from training studies, which "allow for the manipulation of different factors that may contribute to the acquisition of expertise, providing better control over variables influencing this process" and "also allow for better manipulations of the factors that lead to more or less generalization" [28]. Generalization, often discussed as learning transfer, is at the heart of our objective for this study. Although connecting these experiments to the classroom may be a long process, the ultimate goal is to create measures that help us determine if a course is successful in helping students gain perceptual expertise, both relevant and generalizable, in that field.

1.1 Basis for Methods

As our methods are largely taken from cognitive psychology, we detail the studies that form the basis of our methodology. Tanaka, Curran and Sheinberg [27] was one of the first studies to use the experimental paradigm we emulate. Earlier research [30] established that visual experts are able to go directly to subordinate level identifications, bypassing the basic levels novices go through. Basic categories are designated by that level of category in which objects have the most attributes in common and the discontinuity between categories is at its greatest [31]. Categories above this level become broader and often less useful (superordinate) and below this level (subordinate) categories become more specific, with more similarities with other subordinate groups than dissimilarities.

For example, a basic level of a stimulus might be *whale*, with the subordinate level being *Humpback Whale* or *Sperm Whale*, and a basic level *dog* would have subordinate levels such as *poodle* or *beagle.* Both stimuli would share a superordinate category of *animal.* In Tanaka et al. (2005)'s study, researchers asked "to what extent does subordinate-level learning contribute to the transfer of perceptual expertise to novel exemplars and novel categories?" In other words, if experts can identify an object in a particular image, are they also better at identifying new images of that object and new categories of related objects? Tanaka et al. [27] used images of birds: 10 owl species and 10 wading bird species. In this case, *wading bird* and *owl* are the basic categories, with individual species making up the subcategories. Subjects were trained over six days, with final testing on the seventh day. Testing was conducted using a *sequential matching task*, where the subject is shown two images, one after the other, and must respond whether the birds shown are the same or different species. Importantly, some trials used different images of the same species, so subjects needed to identify the species in each image rather than merely matching identical images. One of the training tasks was a *categorization task*, where the subject is shown a single image and must indicate the correct category for the image.

Tanaka et al. (2005) discovered that, for example, novice subjects trained at the subordinate level (on individual owl species) could not only identify new images of the owl species they were familiar with, but the subjects were also able to learn new owl species more quickly. That is, the subjects' training did indeed *generalize* to novel exemplars and to novel categories. However, when learning at the basic level (wading birds), subjects did not demonstrate any generalization, and there was no improvement on basic level identification response times. The subjects had to *perform the task* of noticing differences and categorizing the species in order to gain the perceptual expertise. Mere exposure to all the images was not enough.

The results from the Tanaka et al. 2005 study were replicated by Scott, Tanaka, Sheinberg, and Curran [26]. This study, also using bird images, included the important addition of looking at subjects' brain responses to stimuli, measured in the form of event-related potentials (ERPs), differentiating the brain processes involved with categorizing basic vs. subordinate stimuli. This revealed that different brain mechanisms are activated for these two levels of task, These findings from the 2006 study were then replicated and extended by Scott and her colleagues [25]. Scott, Tanaka, and Sheinberg [25], which used images of cars, not only confirmed that different brain mechanisms were activated for basic and subordinate level tasks, but also included a final assessment after a one-week delay. The data revealed subjects trained at the subordinate level had improved performance, even after a one-week delay, whereas basic level training did not.

2. Methods

Following Tanaka, Curran, and Sheinberg [27], we created a perceptual expertise experiment. After initial testing [32], we chose to use static images of fluid flows that could be sorted as laminar or turbulent in a single session experiment, testing two kinds of participants: "novices" with no prior technical knowledge of fluids, and relative "experts" who had passed at least one college-level fluids course.

2.1 Participants

Individuals who participated in the study were ages 18–30, with normal or corrected-to-normal vision. They gave informed consent to participate in the study, as per the protocol approved by the Institutional Review Board. Participants were recruited via fliers, email, and classroom announcements. The experiment recruited self-reported novices in fluid dynamics (n = 57) and relative "fluids experts" (n = 39) by recruiting students who had completed fluids courses. Most novice participants were recruited through the psychology department website, which gives course credit to students for participating. Most expert participants were recruited through announcements in upper-division engineering courses, populated by students who have already passed a fluids course. All participants trained individually and were given a small cash payment ($10) for their participation. Full demographics reported in Appendix A.

2.2 Materials

This experiment was programmed in MATLAB (version R2013b, The MathWorks, Inc., Natick, MA) using a locally-developed experimental framework (code available at https://github.com/warmlogic/expertTrain) and presented with Psychtoolbox, an open source set of functions for vision and neuroscience research [33, 34]. Using this toolkit, the experiment was presented on a computer, limiting what keys or other controls the participant could use. Participants viewed the experiment on 17-inch flat-panel displays with a resolution of 1024 × 768 (60 Hz frame rate) approximately 45 centimeters in front of the participants, and used a standard QWERTY keyboard.

2.3 Image Selection and Processing

For this experiment, the categories of images were turbulent and laminar fluid flows collected from online sources. To verify that images were classified correctly as either turbulent or laminar, we asked two mechanical engineering professors who regularly teach fluids to classify the images independently. If an image was in doubt or labeled "transitional" by either professor, the image was not included in the study. One specific type of image used was of Von Kármán vortex streets, which are seen when a fluid flows past an obstacle, and the wake becomes a series of vortices. Vortex streets can be either turbulent or laminar (see images a. and b. in Fig. 1). Twenty images of each category were included as stimuli for the experiment for a total of 40 vortex street images. Another group of images (called "general") contained a wide variety of flows, both liquids and gases, none of which were vortex streets. These were also categorized as either lami-

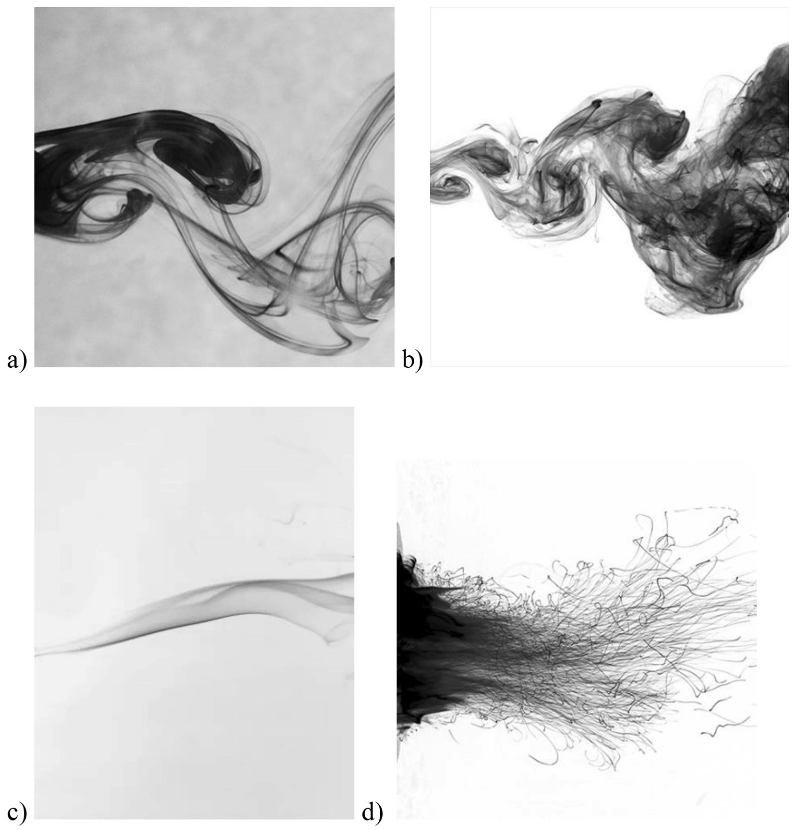

Fig. 1. Processed image examples, (a) laminar vortex street (b) turbulent vortex street (c) laminar general flow (d) turbulent general flow.

nar or turbulent, and likewise, 20 images from each category were included. Thus, our four groups of images were:

1. Laminar instances of Von Kármán vortex streets.
2. Turbulent instances of Von Kármán vortex streets.
3. Laminar instances of general flows (all non-vortex streets).
4. Turbulent instances of general flows (all non-vortex streets).

See Fig. 1 for examples from each group of images.

Several steps were taken to remove extraneous visual information from the images. All images were processed to be gray-scale, no larger than 450×450 pixels, and no less than 230×230 pixels. All vortex street images were oriented with the flow going from left to right. All portions of the display not covered in images or text during the experiment were presented as gray pixels.

2.4 Experiment Design

Participants received instructions that covered what to expect from the format of the experiment, but no information regarding the nature of the images or what the categories would be. The experiment was conducted in a single session, with a typical participant taking 15 minutes. Within the text of the experiment, the categories were called 1 and 2 to reduce bias due to participants' prior familiarity with the words "turbulent" and "laminar." During training, participants learned which images went into each category by trial and error.

Additionally, the categories were randomly assigned to the two response keys for each participant, to reduce bias from hand-dominance. Specifically, keys 'F' and 'J' on a QWERTY keyboard were used as response keys. Because most participants used their left index fingers to press 'F' and right index fingers to press 'J', randomly assigning which response went with which key by participant reduced bias possibly introduced by hand-dominance.

Two different types of task were used:

- Matching task: used as the Testing task. Participants were shown two images, sequentially, and required to indicate whether the two images were in the same or different categories. A fixation point (a black '+') appeared in the center of the screen before each trial. Participants were shown each image for 800 ms with an interstimulus interval ranging from 500–700 ms. Participants were given 2000 ms to respond, while the phrase "Same or Different" is on screen. The sequential matching task was used for all three testing phases (pre-test, post-test, alt-test, more fully described below), and participants received no feedback. See Table 1.
- Categorization task: used as the Training task. Participants were shown a single image, and required to indicate if the image fit category one or two. A fixation point (a black '+') appeared in the center of the screen before each trial. Participants viewed each image for 1000 ms, were given 2000 ms to respond while the phrase "1 or 2" on screen, and were then shown feedback on their choice for 1000 ms. There was a 500 ms pause between trials. For our experiment, the categorization task was used for the single training phase, and participants received feedback for their actions. When correct, participants saw a green-colored "Correct!" and heard a high-pitched beep, and when incorrect, they saw a red-colored "Incorrect" and heard a low-pitched beep.
- A note on image presentation duration: related perceptual expertise studies also use very brief presentation times, such as 2000 ms [35], 1000 ms [36] or even 500 ms [26] with pauses between images ranging from 200 to 1500 ms in those same studies. Initially, we were concerned that the flow images would require longer image presentation times; however, the data from our pilot study of novice participants [32] revealed that these brief image presentation times were sufficient. Although engineers rarely evaluate situations this quickly, keeping presentation duration in line with past perceptual expertise studies facilitated comparison with that prior work.

When participants arrived at the experiment location, they were assigned a number, so that they could be evenly divided for the two training conditions. Even-numbered participants were given the pre-test, training, and post-test on Von Kármán vortex streets. Odd-numbered participants were given general images for the pre-test, training, and post-test.

At the start of the session, participants completed brief practice tasks in matching and categorization to learn the controls for the experiment. Practice images were of everyday objects, and participants were asked to categorize them as solids or liquids. The practice tasks were identical for all participants. After the practice task, the participant had the opportunity to ask questions of the experiment facilitator before continuing onto the actual experiment.

The first test phase, the pre-test, asked participants whether two sequentially presented images matched (a matching task). They were given no guidance as to what criteria to use in attempting this task.

The next phase, the training phase, presented the participants with a categorization task with feedback on each trial. This trained them to sort the flow images as type 1 or 2. This task was deliberately trial-and-error; images were labeled as type 1 or type 2 during feedback to avoid drawing on participants' prior knowledge of laminar and turbulent flows.

The second test phase, the post-test, asked participants to complete a matching task akin to the pre-test. Note that images used in the pre-test, training, and post-test were selected at random from the same set, such that participants were not presented with the same images in each phase.

The final test phase, the alt-test, asked participants to complete a matching task using an alternate set of images. That is, participants trained and tested on Von Kármán vortex streets up to this point were now shown the general flow images, and those trained and tested on the general images were shown the Von Kármán vortex streets. See Table 1 for experiment phases with numbers of trials for each.

At the end of the experiment, participants were asked to write responses for two concept questions, which were set aside for future evaluation:

1. Thinking about your experience in the experiment, how would you describe the two categories of images?
2. How did you decide which images to place in which category?

Lastly, the participants completed a brief demographic survey. The demographic survey was placed at the end to avoid creating stereotype threat during the experiment [37].

During the experiment, half of the images from each category were selected randomly for the training task, while all images were used for the testing tasks. By reserving some of the images, we could determine whether training generalized to the untrained images, decreasing the likelihood that participants were showing improvement solely through memorization of individual images. Note that for the alt-test, all participants were seeing entirely new images.

Both fluids novices and fluids experts were split into these two training groups, such that we had four groups to examine.

2.5 Statistical Procedure

To determine whether participants were able to learn the categories in the brief training task, and then to see if participants were able to transfer that learning to the alternate set of images, we recorded their accuracy during training and each testing phase. Matching task results for each testing phase (pre, post, alt) were tallied. Responses were classified as "hits" (answering "same" when images are from the same category), "misses" (answering "different" when images are from the same category), "correct rejections" (answering "different" when images are from different categories), and "false alarms" (answering "same" when images are from different categories). We then focused on hit rate and false alarms. These two totals indicated a participant's response accuracy, because the number of trials was consistent. For example, the post-test had 40 trials, 20 where "same" was correct, and 20 where "different" was correct. A participant with 14 "hits" would typically have 6 "misses", and if the same participant had 8 "correct rejections," there would be 12 "false alarms."

Originating in electrical engineering, *signal detection theory* is also heavily used in psychology in situations where decisions are made with a degree of uncertainty. One measure commonly used in detection theory is a sensitivity index (d′), in order to separate discrimination from response bias [38]. The sensitivity index (d′) was calculated for each test phase of each subject. This is estimated from the hit and false alarm rates as follows:

$$d' = z(\text{hit rate}) - z(\text{false alarm rate})$$

where $z(x)$, $0 \leq x \leq 1$, is the inverse of the cumulative distribution function of the normal distribution. Because d′ accounts for both the hit rate and the false alarm rate, it allows us to measure participants' ability to differentiate between same and different trials (where "same" was the correct answer and those where "different" was correct), while taking into account response bias for participants who tended to press one key more than the other regardless of the stimuli. Note that $d' = 0$ indicates

Table 1. Phases of experiment with types of images and number of trials for each

Phase (task)	**Pre-test (Phase 1) (Matching)**	**Training (Categorization)**	**Post-test (Phase 2) (Matching)**	**Alt-Test (Phase 3) (Matching)**
V. Street-trained (45 participants)	Vortex Streets (40 trials)	Vortex Streets (20)	Vortex Streets (40)	General Flows (40)
General-trained (47 participants)	General Flows (40 trials)	General Flows (20)	General Flows (40)	Vortex Streets (40)

responses were roughly the same as chance (50% accuracy).

Pre-test sensitivity indices for the four groups were evaluated to verify that the groups were statistically similar prior to training. We performed a one-sided *t*-test for mean d′ change greater than zero for each group for both post-test (post-test d′ – pre-test d′) and alt-test (alt-test d′ – pre-test d′), to assess whether training was effective when images were consistent (post-test) and whether the training generalized (alt-test). All assumptions were confirmed, e.g., we confirmed the normality assumption with the Shapiro-Wilk test.

A two-way ANOVA was performed, using the two training image groups (vortex streets / general images) and two expertise levels (expert / novice). This was performed for the difference in d′ between the pre- and post-tests (post-test d′ - pre-test d′). All analysis was done using R (Version 3.2.3 (2015–12–10)). All ANOVA assumptions were checked and satisfied. A 5% significance level was used for all conclusions.

The concept question responses and demographic data were logged for future study.

2.6 Data Preparation

Data from three participants (one novice, two experts) became corrupted and were removed from the analysis. Thus, data from a total of 56 novice participants and 37 expert participants were analyzed. We looked for outliers whose performance on any one test phase was more than three standard deviations away from the mean score, and removed one expert participant from the data as a consequence. That participant's data revealed a very high false alarm (false positive) rate. In fact, the participant responded with the same key for 39 of the 40 trials in the alternate testing phase. The analysis was conducted for the remaining participants:

1. Novices trained on vortex streets (n = 28) sometimes referred to as Novices (vortex)
2. Novices trained on general flows (n = 28) sometimes referred to as Novices (general)
3. Experts trained on vortex streets (n = 17) sometimes referred to as Experts (vortex)
4. Experts trained on general flows (n = 19) sometimes referred to as Experts (general).

We then calculated sensitivity gain for each subject. Sensitivity gains were calculated as: (d′ on post-test – d′ on pre-test).

3. Results

The mean (standard deviation) value of the sensitivity index (d′) at each of the three phases of the study is provided in Table 2. Fig. 2 presents the same three mean values per group, plotted with bolded diamonds and connected with solid lines to indicate trends. In addition, Fig. 2 shows the individual sensitivity index values for each participant; these values are connected with dashed lines. Both Table 2 and Fig. 2 show a general pattern of improvement from pre- to post-test and a retrogression of these improvements between the post- and alt-test. These summary observations will be discussed in more formal statistical detail next.

Pre-test sensitivity indices (d′) were evaluated for each of the four groups in order to verify that the groups were statistically comparable at the outset. Table 2 reveals how all but the group of novices trained on general images averaged a d′ between 0.45 and 0.49. The Novice (general) group's mean pre-test sensitivity was lower, at 0.25. The ANOVA for any between group differences in mean pre-test sensitivity yielded $p = 0.25$ ($F(3,88) = 1.40$). Therefore, there was not statistical evidence of any differences in the means of the four study groups in the pre-test. This confirmed there were no systematic group differences in d′ prior to training; thus change score analysis using ANOVA was appropriate for comparisons of post to pre d′ values.

We proceeded with comparisons of the three test phases: pre-test, post-test, and alt-test. Table 3 summarizes the change in d′ from the pre- to post-test by group; that is, Table 3 details the per-group training effect. To be clear, the change in d′ here is found by subtracting the pre-test d′ from the post-test d′. Each group showed a statistically significant positive change when we performed a one-sided *t*-test. Because the pre-test data was tested twice, the *p*-values have been adjusted for multiple compar-

Table 2. Mean and standard deviation of d′ for each group in each test phase

Group	Pre-test (Phase 1) Mean (standard deviation)	Post-test (Phase 2) Mean (standard deviation)	Alt-test (Phase 3) Mean (standard deviation)
Experts trained on Vortex Streets	0.49 (0.62)	1.47 (0.79)	0.88 (0.72)
Novices trained on Vortex Streets	0.46 (0.41)	1.07 (0.70)	0.33 (0.48)
Experts trained on General Images	0.45 (0.38)	1.24 (0.84)	0.71 (0.57)
Novices trained on General Images	0.25 (0.42)	0.72 (0.56)	0.68 (0.62)

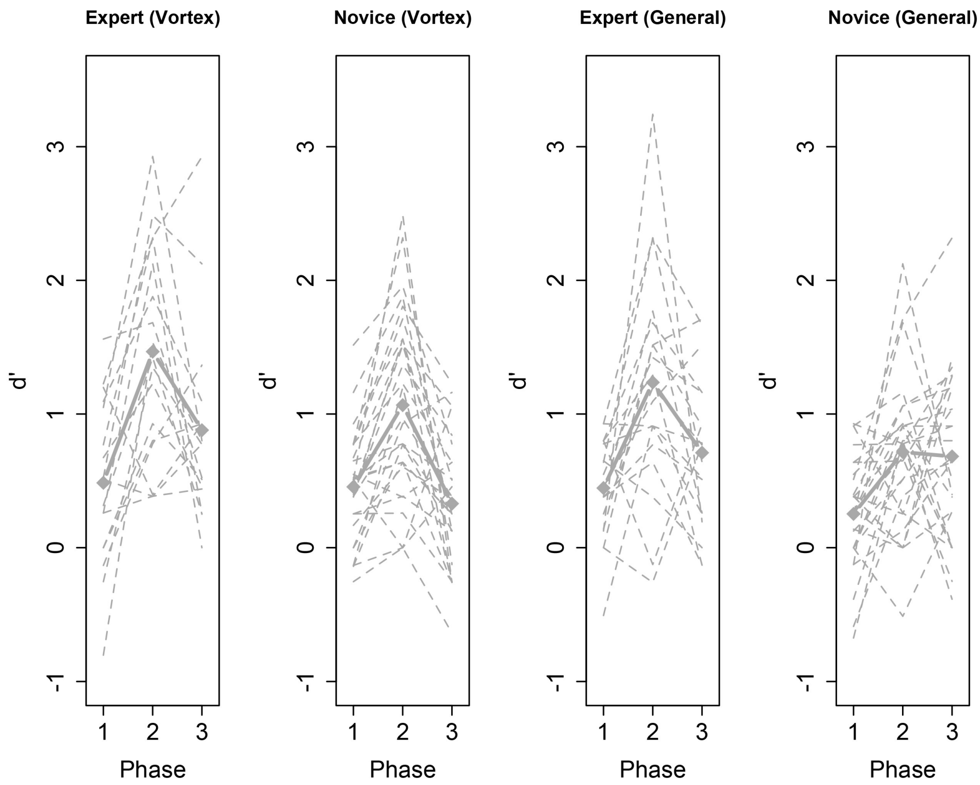

Fig. 2. Sensitivity indices (d′) for each group by test phase. Bolded line indicates the mean. Individual participant values are in dotted lines. Phase 1 – pre-test when participants were completely guessing, Phase 2 – post-test occurred after a training phase, Phase 3 – alt-test, when an alternate set of images were used.

isons. More, only one group (Novices (general)) did not have at least 75% of its participants with improved d′ values. To illustrate this further, a side-by-side boxplot for the change in d′ values was created (Fig. 3) with a horizontal dashed line at 0 to denote no change. Most individual participants showed improvement (positive delta), suggesting a positive effect from the training phase.

For pre-to post-test comparisons, the summary statistics (Table 3) and boxplots (Fig. 3) display evidence that the experts improved more relative to novices on both training image types. This, then, is a key finding in this study: the difference between experts and novices, when training image type was taken into account in the two-way ANOVA, resulted in a statistically significant difference of –0.3485 for novices versus experts (p = 0.0266, 95% CI [-0.6557, -0.0413]). In comparison, there was no statistically significant difference when examining those trained on vortex streets versus those trained on general images, (p = 0.2758, mean of 0.1655, 95% CI [-0.1344, 0.4654]). An interaction effect between expertise and training image type was not significant (p = 0.8939) and therefore not included in the final model. Thus experts learned to categorize laminar and turbulent flows more accurately than novices.

The last part of the statistical analysis was to compare the change in sensitivity index from the alt- to pre-test. A central question of the present study was whether participants would *generalize* their learning from the images they were trained on to a new set; that is, would participants trained on the general images be able to match laminar and turbulent for vortex street images and would participants trained on vortex streets be able to match laminar

Table 3. The mean (and standard deviation) and median of the change in participants' sensitivity indices from post-test – pre-test. p-values here indicate that the change was non-zero, showing that training was effective for all four groups

Group	Mean (Standard Deviation)	Median	*p*-value adjusted for multiple comparisons
Experts trained on Vortex Streets	0.98 (0.84)	0.93	p = 0.0002
Novices trained on Vortex Streets	0.61 (0.55)	0.61	p < 0.0001
Experts trained on General Images	0.79 (0.91)	0.87	p = 0.0014
Novices trained on General Images	0.46 (0.67)	0.39	p = 0.0011

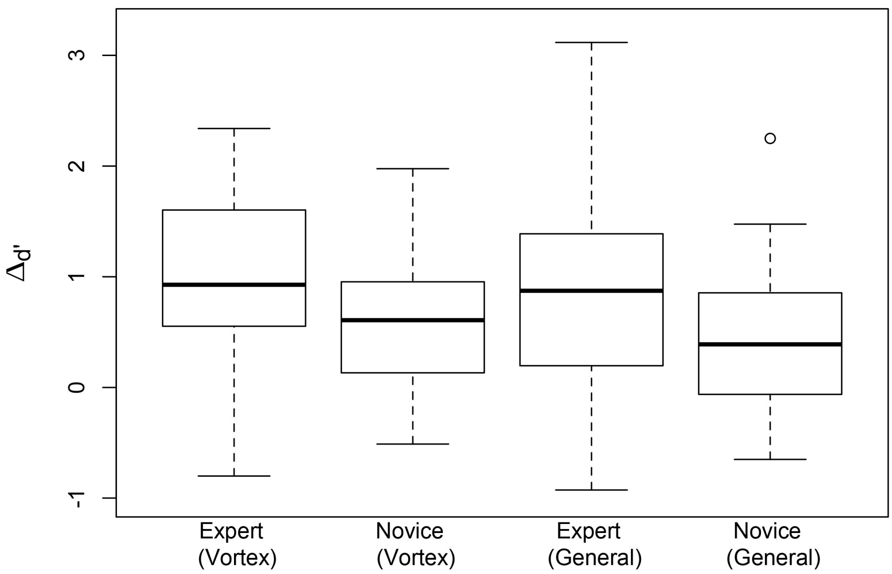

Fig. 3. Changes in sensitivity index (d′) by participant group, comparing the post-test to the pre-test. Bold lines indicate the median, boxes cover the 25th to 75th percentile, and whiskers denote the maximum and minimum values. Positive values indicate increased sensitivity to detecting change, interpreted as improvement after training.

and turbulent when comparing general images. Furthermore, would these alt- to pre-test comparisons vary by expertise level (expert/novice)?

Table 4 summarizes the change in d′ from the pre- to alt-test by group where the change is found by subtracting the pretest d′ from the alt-test d′. We performed a one-sided t-test for mean d′ change greater than zero for each group for alt-test d′ – pre-test d′. Because pre-test data was tested twice, the *p*-values reported in Table 4 have been adjusted to account for multiple comparisons. A side-by-side boxplot for the median change in d′ values was created (Fig. 4) with a horizontal dashed line added to reflect when the change equaled 0. Notice that the Novices (general) show statistically significant improvement, whereas the results for novices trained on vortex streets are as if they had received no training; $p = 1.0$. This is an indicator that generalization to alt images was most difficult for novices who trained on vortex street images and were transferred to general images. This aligns with results from prior perceptual expertise studies.

The results for the expert participants are less clear. See Table 4. They show a general improvement over the pre-test. Yet, the necessity of doubling *p* values (because the pre-test data is used twice) results in $p > 0.05$ for both Experts (vortex streets) ($p = 0.0755$) and Experts (general) ($p = 0.1168$). In addition, although the median of Experts (vortex streets) improvement is near zero, the mean is 0.39, suggesting that those who did improve made large gains, as seen in Fig. 2.

4. Discussion

Our study began with anecdotes from students about the moments when they became aware of fluid physics in everyday life. We are motivated by a need to understand the connection between perceptual expertise of fluid flows and students' conceptual understanding. Unlike previous perceptual expertise studies, which focused on images of birds or cars, our study asks participants to sort fluid flow images by somewhat more abstract characteristics: laminar and turbulent. Our first question was whether participants, particularly novices, would be able to learn to sort these stimuli in a similar fashion as participants in past studies. (The null hypothesis for this question is that training would have no effect, and we would see no real difference

Table 4. The mean effect (with 95% confidence intervals) of the change in participants' sensitivity indices from alt-test – pretest

Group	Mean (95% CI)	*p*-value (adjusted for multiple comparisons)
Experts trained on Vortex Streets	0.39 (–0.0453, 0.8327)	0.0755
Novices trained on Vortex Streets	–0.13 (–0.3399, 0.0890)	1.0000
Experts trained on General Images	0.26 (–0.0727, 0.6014)	0.1168
Novices trained on General Images	0.43 (0.1638, 0.6971)	0.0026

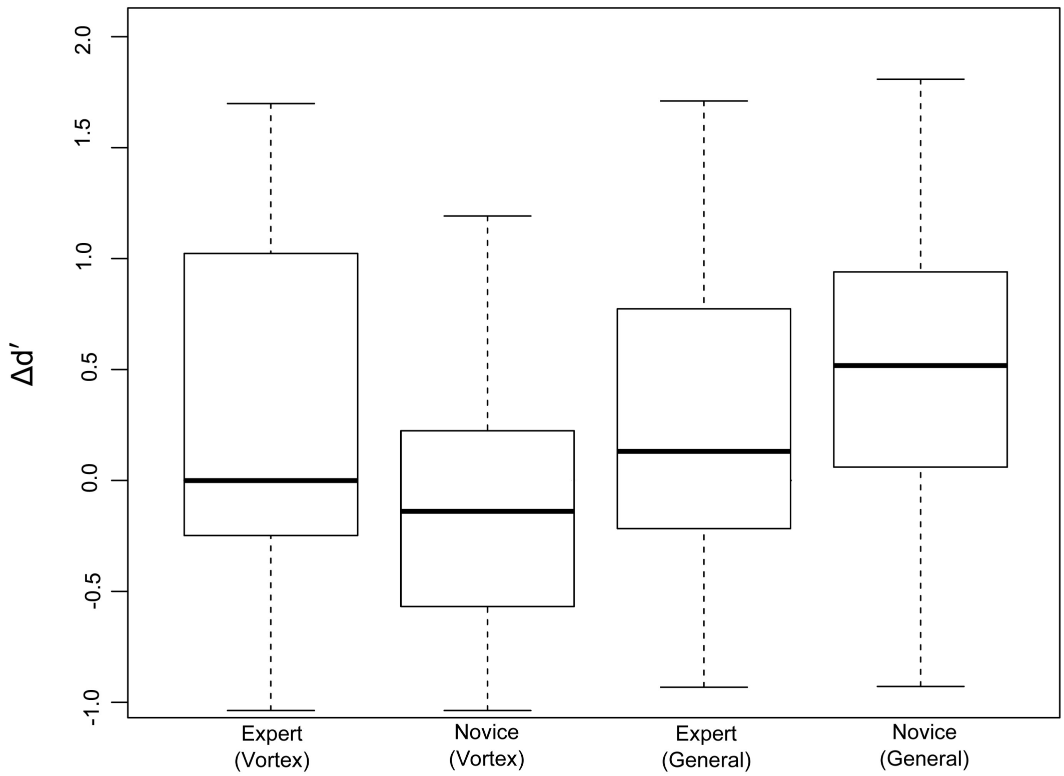

Fig. 4. Change in sensitivity index (d′) by participant group, comparing the alt-test to the pre-test. Median values are indicated with bold lines; with boxes span 25th to 75th percentile and whisker denote the maximum and minimum values. Positive values indicate increased sensitivity to detecting change, interpreted as learning generalizing to the new image type.

between the pre- and post-tests.) We found that initial performance (pre-test) did not vary across groups, so expertise did not influence participants' initial categorization abilities. However, expertise did benefit participants' ability to learn, as demonstrated by their greater gain between post-test and pre-test. If prior knowledge of fluids made no difference, we would have expected a significant difference between all of those trained on vortex streets when compared to all of those trained on the general images, and no difference when comparing novices to experts.

A second question was whether our study mirrored findings in prior studies [39], which demonstrated that training on specific instances (in our case, vortex streets) results in greater learning (improved sensitivity) for those same stimuli, yet that training does not generalize well. (The null hypothesis for this question is that image type would have no difference that participants trained on vortex streets would have similar results to those trained on the general images.) That earlier work found that novices training with a more variable group of exemplars (like our general flows) resulted in less improvement in the post-test, but that sensitivity is more likely to persist when presented with the new set of exemplars [28, 40, 41]. Our results for novices mirror this finding remarkably well. Novices trained on general images show statistically significant improved sensitivity when tested on vortex streets, but novices trained on vortex streets did not show any change in sensitivity to general images. See Table 4.

Since novices appear to learn fluid flow stimuli in a similar fashion as novices learning car or bird images, the expert participant results become even more interesting. Our third question was whether experts and novices would perform the same in the study. (The null hypothesis for this question is that there would be no difference between the groups.) The expert participants did not exhibit the same pattern of learning (sensitivity gain) from general to specific categories of images, or vice versa, as novices. On the contrary, there was a positive effect from having expertise in fluids that was not dependent on training image type. The trend in the data suggests that the general or specific nature of the training stimuli seems to matter for the novices, but not for the experts. The experts' results are admittedly not statistically significant, however the contrast between novices and experts implies that expert participants are displaying a type of knowledge transfer.

Some may argue that this result is not noteworthy when taken in full context. After all, participants were alerted to the involvement of fluids content because recruiting efforts mentioned the need for participants with and without that experience. Likewise, participants were asked whether they had fluids knowledge as they signed in for the experiment. Future studies may want to alter recruitment methods in order to avoid mentioning the content area of the stimuli prior to testing. One approach might be to recruit solely from engineering courses that have Fluid Mechanics as a prerequisite, which would eliminate the need to explicitly mention fluids. A questionnaire after testing could confirm this precondition, so that the question would not prime the participants for a fluids-oriented task beforehand.

Even noting this concern of priming participants to expect something fluids-related, this task is probably unlike those they have performed in other contexts. The work required for fluid mechanics courses tends to be analytical, focused on solving equations. The use of images in the course is typically limited. Some students may have used drawings to aid their comprehension, and some professors expose students to images of fluid flows and discuss these images in class. In fact, the fluids courses taken by our "expert" students often included a "Flow Vis of the Day." Even with exposure to images, short video, and computer simulations of fluids, students rarely had to complete tasks involving analysis of images of fluid flows. The exception would be students who had participated in a course explicitly about creating and understanding such images, called Flow Visualization [22]. Unfortunately, we did not ask participants whether they had participated in this class at the time of the experiment.

From Tanaka, Curran and Sheinberg [27], we have evidence in a laboratory setting that exposure to images is not sufficient for novice participants to learn to distinguish between stimuli categories at the subordinate level. They found participants had to perform the sorting tasks in order to gain the perceptual expertise. In our case, expert participants were applying what they knew prior to the experiment to a new task. This study is not directly analogous to studies that engaged bird and car experts. Bird watchers, in particular, commonly use their expertise in visual tasks. In contrast, the typical fluid mechanics student is not trained to perform visual tasks.

What our study does suggest is a crossover between conceptual and visual perception skillsets. Conceptual knowledge learned in the classroom influenced their performance on a subsequent perceptual task. We view the ability to map conceptual understanding onto visual information as a necessary step toward the expansion of perception that is part of the transformative experience.

We know that analytical problem-solving does not imply the ability to handle a problem presented visually. Work in the Physics Education Research (PER) community has documented that students who can correctly use Ohm's law and Kirchhoff's rules to solve complicated quantitative circuit problems often have difficulty with a simpler qualitative task ranking light bulbs in a circuit diagram by brightness [42]. In the PER study, only 15% of the students showed that they can transfer an analytical skill to a more qualitative task. Perhaps what is missing is an underlying conceptual understanding. If handling ideas visually promotes connections between analytical skills (solving equations) and conceptual understanding, then that would be another reason to implement visual tasks in STEM courses.

In our preliminary review of the open-ended questions, 27 of the expert participants identified their task as sorting by laminar and turbulent; only two of the novice participants did so [43]. Of the expert participants who did not identify the task correctly, one believed the task was sorting real photos from screen captures of computer generated simulations, and five believed the task was sorting by phase (gases vs. liquids). Recognition that students may doubt what they have learned when confronted with new data or situations is worth further investigation, and may be related to other ways of framing the transfer problem, such as "activating resources" [4]. We would also like to identify which features of the training phase activated their knowledge of laminar and turbulent flows, such that expert participants knew to apply something learned in class to this new, visual matching task. We posit that this process may be similar to the analogical scaffolding proposed by Podolefesky and Finkelstein [44], wherein learning is bidirectional between an analogy used to teach a concept and the target concept. In our case, learning may be bidirectional between the visual task and the analytical tasks from their coursework, such as solving equations.

The main contribution of this study is that it begins to create a linkage between conceptual learning and visual expertise, with the potential to create an alternate assessment of conceptual understanding separate from analytical problem-solving. This study establishes a viable method of measuring visual perceptual expertise in a specific dimension of an engineering discipline.

4.1 Limitations

Limitations to this study include a small sample size. We began by running the experiment in the Psychology Department building, but eventually realized that the specialized engineering population we needed to recruit simply would not walk the fifteen minutes to a different building to participate. Participation increased once the experiment was moved to the Engineering Center. Another limitation was the scope of the experiment itself: we only tested one visual dimension of fluids – laminar versus turbulent. A challenge with creating more dimensions for the experiment was locating and processing suitable images. Increasing both the size of the study and the range of characteristics tested would greatly improve further work.

4.2 Future Work

We are excited by the possibilities for the perceptual expertise study; there are many options for extending this work. First, we plan to analyze the open-ended concept questions. We have begun qualitative coding of these concept question responses, for both expert and novice participants, and other meaningful trends may emerge.

As we adapt the experiment for additional participants, we may include other concepts from fluids, including jets, shear layers, and Rayleigh-Taylor instabilities, in order to create a more multi-dimensional measure of visual expertise in fluids. This could potentially become a useful classroom assessment.

We would also like to switch to a web-based platform so that other institutions could more easily join the study. We are already looking at potential platforms. Once the experiment can be more widely distributed, there are more options for establishing its relevance. We imagine, for instance, asking fluids experts from around the world to participate with a larger group of images, to create a more definitive range of fluids expertise, not unlike what past studies have measured in bird watchers. This work may also inform studies of the brain, relating back to the other studies of perceptual expertise cited earlier. Are the abstract categories of fluid flows utilizing different brain mechanisms than more concrete stimuli, like cars or faces?

Other future work could continue to explore the connections between visual expertise and conceptual understanding in other disciplines, such as structural analysis, failure analysis, material microscopy, and even medical diagnoses. Such work may lead directly to conceptual assessment, and will help further characterize the connection between perceptual expertise and expanded perception as defined in the transformative experience. As such, this work represents another collaboration between current cognitive psychology and discipline-based education research. These kinds of collaborations are gaining attention as more researchers attempt to bridge the disciplinary divide [45, 46].

This is not a call for merely inserting more images into fluids courses. We know from prior studies that exposure to images is not enough for participants to gain the perceptual expertise needed to sort those images at the subordinate level [27]. Similarly, mere exposure to flow images without instruction would likely be ineffective. Fascinating images may attract some students to study fluids, but ideally these images would also draw the students' curiosity about constraint-based interaction (CBI) causing a particular image. Even limited use, as is the case in some fluid mechanics courses, of images, videos, and computer simulations of fluid movement, collectively known as flow visualization or "flow vis", helps cement understanding. As one student in a related study commented, "I think I kind of would have got it without the visualization, but I think the visualization really locked in the concepts and the knowledge" [47]. So, while we encourage the thoughtful use of images within lessons to reinforce understanding for students who otherwise might not understand the concepts, we believe it is even more important to devise learning activities where students create, manipulate, or describe images related to central concepts. These activities may provide students with an additional means of relating to the content, as well as provide instructors with an additional means to gauge comprehension. The possibility of utilizing visual perception as a means of boosting learning in fluid dynamics suggests the benefit of including other perceptual expertise. As a result, we are exploring the possibilities of not only visual, but also auditory and tactile cues that might enhance learning in engineering courses.

5. Conclusions

Our perceptual expertise experiment with images of laminar and turbulent flow suggests (1) that novices learned the concepts used to sort the flow stimuli in ways similar to participants in prior studies, which used more concrete stimuli, and (2) that the participants with prior fluids knowledge (experts) did significantly better than the novices, regardless of the images used in training. This result suggests that these participants were able to access their conceptual knowledge about fluids to perform this new, visual task: sorting the images by whether the images were of laminar or turbulent flows. This idea, while seemingly simple, opens the door to new ways of understanding conceptual learning, and it causes us to question whether this interaction is two-way. That is, for the novices who learn this visual perception task, would learning the concepts of fluid physics around laminar and turbulent be easier as a result? Also, could we use such a task as a type of assessment in a fluids course, using visual expertise in fluids images as one benchmark of learning? Experiments like this one could assist in differentiating students who both understand the concepts and can work through the mathematical procedures from those who can complete the math but do not grasp the underlying concepts.

Acknowledgements – This material is based upon work supported by the National Science Foundation under Grant No. EC-1240294 and supported in part by the Chancellor's Award from the Center for STEM Learning, University of Colorado Boulder. All photos in Fig. 1. are used with artists' permission. Image (a) by Andrew A. Van Der Volgen, Jonathan Cook, Finn Ostrem, Kyle Samples, 2015. Image (b) by Kyle Samples, Jonathan Cook, Andrew Van Der Voglen, and Finn Ostrem, 2015. Image (c) by Hunter Miller, 2016. Image (d) by Marco Gardi, 2016. Thanks to Garrison Vigil for assistance with data collection and image processing. Thanks to Krystin Corby, Shaina Martis, and other assistants of the Curran lab for guiding many of the participants through the experiment.

References

1. R. D. Pea, Socializing the knowledge transfer problem, *Int. J. Educ. Res.*, **11(**6), pp. 639–663, 1987.
2. C. M. Bukach, W. S. Phillips and I. Gauthier, Limits of generalization between categories and implications for theories of category specificity, *Atten. Percept. Psychophys.*, **72**(7), pp. 1865–1874, 2010.
3. M. Rodd, Geometrical Visualisation–Epistemic and Emotional, *Learn. Math.*, **30**, 2010.
4. D. Hammer, A. Elby, R. E. Scherr and E. F. Redish, Resources, framing, and transfer, *Transf. Learn. Mod. Multidiscip. Perspect.*, pp. 89–120, 2005.
5. J. S. Brown, A. Collins and P. Duguid, Situated Cognition and the Culture of Learning, *Educ. Res.*, **18**, pp. 32–42, 1989.
6. L. C. Benson, M. K. Orr, S. B. Biggers, W. F. Moss, M. W. Ohland and S. D. Schiff, Student-Centered Active , Cooperative Learning in Engineering, *Int. J. Eng. Educ.*, **26**(5), pp. 1097–1110, 2010.
7. Y. Engeström, Non Scolae Sed Vitae Discimus: Toward Overcoming the Encapsulation of School Learning, *Learn. Instr.*, **1**, pp. 243–259, 1991.
8. L. R. Lattuca, P. T. Terenzini, J. F. Volkwein and D. Peterson, George, The Changing Face of Engineering Education, *The Bridge*, **36**, pp. 5–13, 2006.
9. B. D. Lutz, S. A. Brown and N. Perova-Mello, Exploring Practicing Engineers' Understanding of Fluid Mechanics Concepts, *Int. J. Eng. Educ.*, **35(**2), pp. 535–547, 2019.
10. R. J. Spiro, P. J. Feltovich, M. J. Jacobson and R. L. Coulson, Cognitive Flexibility , Constructivism , and Hypertext: Random Access Instruction for Advanced Knowledge Acquisition in Ill-Structured Domains, in T. M. Duffy and D. H. Jonassen (eds), *Constructivism and the Technology of Instruction: A Conversation*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 57–75, 1992.
11. M. T. H. Chi, Creativity: Shifting Across Ontological Categories Flexibly, in T. B. Ward and S. M. Smith (eds), *Creative thought: An investigation of conceptual structures and processes*, American Psychological Association, Washington, DC, pp. 209–234, 1997.
12. R. J. Spiro, B. P. Collins, J. J. Thota and P. J. Feltovich, Cognitive Flexibility Theory: Hypermedia for Complex Learning, Adaptive Knowledge Application, and Experience Acceleration, *Educ. Technol.*, Sept-Oct, pp. 5–10, 2003.
13. C. Bastin, D. Feyers, S. Majerus, E. Balteau, C. Degueldre, A. Luxen, P. Maquet, E. Salmon and F. Collette, The neural substrates of memory suppression: a FMRI exploration of directed forgetting, *PloS One*, **7**(1), p. e29905, 2012.
14. W. Hirst, E. A. Phelps, R. L. Buckner, A. E. Budson, A. Cuc, J. D. E. Gabrieli, M. K. Johnson, K. B. Lyle, C. Lustig, M. Mather, R. Meksin, K. J. Mitchell, K. N. Ochsner, D. L. Schacter, J. S. Simons and C. J. Vaidya, Long-term memory for the terrorist attack of September 11: Flashbulb memories, event memories, and the factors that influence their retention, *J. Exp. Psychol. Gen.*, **138**(2), pp. 161–176, 2009.
15. K. J. Pugh, Transformative Experience: An Integrative Construct in the Spirit of Deweyan Pragmatism, *Educ. Psychol.*, **46**(2), pp. 107–121, 2011.
16. J. Dewey, *Art as Experience*, Perigee by Penguin Group, New York, 1934.
17. J. Dewey, *Experience and Education*, Touchstone by Simon & Schuster, New York, 1938.
18. K. J. Pugh and M. Girod, Science, Art, and Experience: Constructing a Science Pedagogy From Dewey's Aesthetics, *J. Sci. Teach. Educ.*, **18**(1), pp. 9–27, 2007.
19. K. J. Pugh, L. Linnenbrink-Garcia, K. L. K. Koskey, V. C. Stewart and C. Manzey, Motivation, learning, and transformative experience: A study of deep engagement in science, *Sci. Educ.*, **94(**1), pp. 1–28, 2010.
20. B. C. Heddy and G. M. Sinatra, Transforming Misconceptions: Using Transformative Experience to Promote Positive Affect and Conceptual Change in Students Learning About Biological Evolution, *Sci. Educ.*, **97**(5), pp. 723–744, 2013.
21. J. K. Garner, A. Kaplan and K. Pugh, Museums as Contexts for Transformative Experiences and Identity Exploration. Museums as Contexts for Transformative Experiences and Identity Development, *J. Mus. Educ.*, **41**(4), pp. 341–352, 2016.
22. J. Hertzberg, B. Leppek and K. Gray, Art for the Sake of Improving Attitudes towards Engineering. AC 2012-5064, in *ASEE Conference Proceedings*, San Antonio, TX., 2012.
23. I. Gauthier, T. Curran, K. M. Curby, and D. Collins, Perceptual interference supports a non-modular account of face processing, *Nat. Neurosci.*, **6(**4), pp. 428–432, 2003.
24. J. W. Tanaka and T. Curran, A neural basis for expert object recognition, *Psychol. Sci.*, **12**(1), pp. 43–47, 2001.
25. L. S. Scott, J. W. Tanaka, D. L. Sheinberg and T. Curran, The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study, *Brain Res.*, **1210**, pp. 204–215, 2008.
26. L. S. Scott, J. W. Tanaka, D. L. Sheinberg and T. Curran, A reevaluation of the electrophysiological correlates of expert object processing, *J. Cogn. Neurosci.*, **18**(9), pp. 1453–1465, 2006.
27. J. W. Tanaka, T. Curran and D. L. Sheinberg, The training and transfer of real world perceptual expertise, *Psychol. Sci.*, **16**(2), pp. 145–151, 2005.
28. L. S. Scott, J. W. Tanaka, T. Curran, I. E. Gauthier, M. J. E. Tarr and D. E. Bub, Degrees of expertise., in I. Gauthier, M. J. Tarr, and D. Bub, (eds), *Perceptual Expertise Bridging Brain and Behavior*, Oxford University Press, Oxford, pp. 107–138, 2009.

29. L. Vygotsky, *Thought and Language*, The MIT Press, Cambridge, Mass, 1934.
30. J. W. Tanaka and M. Taylor, Object categories and expertise: Is the basic level in the eye of the beholder?, *Cognit. Psychol.*, **23**(3), pp. 457–82, 1991.
31. E. Rosch, Principles of categorization, in D. J. Levitin, (ed), *Foundations of cognitive psychology: Core readings*, The MIT Press, Cambridge, Mass, pp. 251–270, 2002.
32. K. Goodman, J. Hertzberg, T. Curran and N. Finkelstein, Expanding Perception: How Students 'See' Fluids, in *ASEE Conference Proceedings*, Seattle, WA, United states, pp. 26.713.1–26.713.10, 2015.
33. D. H. Brainard, The Psychophysics Toolbox, *Spat. Vis.*, **10**, pp. 433–436, 1997.
34. Psychtoolbox-3 – Overview [Online], http://psychtoolbox.org/, Accessed: 30-Sep-2019.
35. G. Herzmann, V. Willenbockel, J. W. Tanaka and T. Curran, The neural correlates of memory encoding and recognition for own-race and other-race faces, *Neuropsychologia*, **49**(11), pp. 3103–15, 2011.
36. G. Herzmann and T. Curran, Experts' memory: An ERP study of perceptual expertise effects on encoding and recognition, *Mem. Cognit.*, **39**, pp. 412–432, 2011.
37. C. M. Steele, S. J. Spencer and J. Aronson, Contending with group image: The psychology of stereotype and identity threat, *Adv. Exp. Soc. Psychol.*, **34**, pp. 379–440, 2002.
38. N. A. Macmillan and C. D. Creelman, *Detection Theory: a user's guide*, 2nd ed. Erlbaum, Mahweh, NJ, 2005.
39. L. K. Perry, L. K. Samuelson, L. M. Malloy and R. N. Schiffer, Learn Locally, Think Globally: Exemplar Variability Supports Higher-Order Generalization and Word Learning, *Psychol. Sci.*, **21**(12), pp. 1894–1902, 2010.
40. J. Loewenstein and D. Gentner, Spatial mapping in preschoolers: close comparisons facilitate far mappings, *J. Cogn. Dev.*, **2**(2), pp. 189–219, 2001.
41. J. B. Childers and J. H. Paik, Korean- and English-speaking children use cross-situational information to learn novel predicate terms, *J. Child Lang.*, **36**(1), pp. 201–24, 2009.
42. L. C. McDermott, Guest Comment: How we teach and how students learn – A mismatch?, *Am. J. Phys.*, **61**(4), p. 295, 1993.
43. K. Goodman and J. Hertzberg, Work in Progress: Qualitative Insights from a Visual Expertise Experiment in Fluid Mechanics: American Society for Engineering Education, presented at the *ASEE Annual Conference and Exposition*, Tampa, FL, 2019.
44. N. S. Podolefsky and N. D. Finkelstein, Analogical scaffolding and the learning of abstract ideas in physics: An example from electromagnetic waves, *Phys. Rev. Spec. Top. – Phys. Educ. Res.*, **3**(1), p. 010109, 2007.
45. M. A. McDaniel, J. P. Mestre, R. F. Frey, R. Gouravajhala, R. C. Hilborn, T. Miyatsu, J. W. Morphew and L. L. Slakey, Maximizing undergraduate STEM learning: Promoting research at the intersection of cognitive psychology and discipline-based education research, *Report of an NSF Conference 2017*, funded by grant DUE 1630128, https://circle.wustl.edu/white-paper-maximizing-undergraduate-stem-learning/, Accessed 30 Sept 2019.
46. J. P. Mestre, A. Cheville and G. L. Herman, Promoting DBER-Cognitive Psychology Collaborations in STEM Education, *J. Eng. Educ.*, **107**(1), pp. 5–10, 2018.
47. K. Goodman, The Transformative Experience in Engineering Education, Ph.D. thesis, University of Colorado, Boulder, CO, 2015.

Appendix A

Participant Demographics

	Novices (n = 56)	Experts (n = 36)
Mean age	20.64	22.29
% Male/female/other	50% / 46% / 4%	64% / 36% / 0%
% White	82.14%	83.33%
% Asian	14.29%	8.33%
% Hispanic/Latino	7.14%	11.11%
% all others / unreported	0.00%	2.78%
% multi-racial*	3.57%	5.56%

* Participants who selected this also selected other identifications, causing totals to be higher than 100%.

Katherine Goodman is an Assistant Professor and the Associate Director of Inworks, an interdisciplinary initiative at University of Colorado Denver, where she teaches human-centered design and introductory data science, among other courses. Her research focuses in engineering education, particularly on "transformative experiences" – when students change their perception of everyday life to include the concepts they have learned in the classroom. Currently, she is researching freshman engineering student persistence and retention, in terms of how they form a STEM identity. This work is part of the Urban STEM Collaboratory, a shared project of CU Denver, University of Memphis, and Indiana University – Purdue University Indianapolis.

Jean Hertzberg is currently an Associate Professor of Mechanical Engineering at CU-Boulder. She teaches graduate and undergraduate courses in measurement techniques, thermodynamics, fluid mechanics, heat transfer, design and computer tools. She has pioneered a spectacular course on the art and physics of flow visualization, and is conducting research on the impact of the course with respect to visual perception and educational outcomes. Her disciplinary research centers around pulsatile, vortex dominated flows with applications in both combustion and bio-fluid dynamics. She is also interested in a

variety of flow field measurement techniques. Current projects include study of velocity and vorticity in human cardiac ventricles and large vessels.

Tim Curran is a Professor of Psychology and Neuroscience at CU Boulder. He teaches graduate and undergraduate courses in cognitive neuroscience and memory. His research has focused on human learning, memory, visual object recognition, and perceptual expertise. This research has included cognitive/behavioral studies, neuropsychological studies in brain-injured patients, electroencephalography (EEG), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI). He is currently Associate Editor for Journal of Experimental Psychology: General.

Erin Austin is an assistant professor in the Department of Mathematical and Statistical Sciences at the University of Colorado Denver. His research focuses on the implementation of classic and development of novel statistical methodology utilizing: penalized regression, longitudinal analysis, survival modeling, cluster analysis, gene-level association testing, and biostatistics. Erin has a strong interest in multi-disciplinary collaborations, especially in the areas of education and epidemiology.