PERGAMON

Vision Research 40 (2000) 97–109

Vision Research

www.elsevier.com/locate/visres

Mechanisms of generalization in perceptual learning

Zili Liu [a,*], Daphna Weinshall [b]

[a] *Department of Psychology, Rutgers University, Newark, NJ 07102, USA*

[b] *Institute of Computer Science, Hebrew University, 91904 Jerusalem, Israel*

Received 17 November 1998; received in revised form 8 June 1999

Abstract

Learning in many visual perceptual tasks has been shown to be specific to practiced stimuli, while new stimuli have to be learned from scratch. Here we demonstrate generalization using a novel paradigm in motion discrimination where learning has been previously shown to be specific. We trained subjects to discriminate directions of moving dots, and verified the previous results that learning does not transfer from a trained direction to a new one. However, by tracking the subjects' performance across time in the new direction, we found that their speed of learning doubled. Therefore, we found generalization in a task previously considered too difficult to generalize. We also replicated, in a second experiment, transfer following training with 'easy' stimuli, when the difference between motion directions is enlarged. In a third experiment we found a new mode of generalization: after mastering the task with an easy stimulus, subjects who have practiced briefly to discriminate the easy stimulus in a new direction generalize to a difficult stimulus in that direction. This generalization depends on both the mastering and the brief practice. The specificity of perceptual learning and the dichotomy between learning of 'easy' versus 'difficult' tasks have been assumed to involve different learning processes at different cortical areas. Here we show how to interpret these results in terms of signal detection theory. With the assumption of limited computational capacity, we obtain the observed phenomena — direct transfer and acceleration of learning — for increasing levels of task difficulty. Human perceptual learning and generalization, therefore, concur with a generic discrimination system. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Perceptual learning; Practiced stimuli; Learned stimuli

1. Introduction

Learning in biological systems is of great importance at all levels of information processing, including perceptual and cognitive learning. But while cognitive learning (or 'problem solving') is typically abrupt and generalizes to analogous problems, perceptual skills appear to be learned gradually and specifically. Thus, human subjects cannot generalize a perceptual discrimination skill to solve similar problems with a different attribute. For example, in a visual discrimination task (Fig. 1), a subject who is trained to discriminate between motion directions 43 and 47° cannot use this skill, 90° away, to discriminate between 133 and 137°.

Many perceptual learning experiments have used the following experimental paradigm: subjects are trained in a discrimination task with one particular attribute (e.g. the pair of motion directions 43 and 47° above), and are later tested with another attribute (e.g. the second pair of motion directions 133 and 137°). Typically subjects start off with poor discrimination, and improve substantially with training. But when tested with the new attribute, their performance drops to baseline. In other words, learning does not transfer from the first to the second attribute.

To give a few examples, Fiorentini and Berardi (1980) have found that learning in waveform discrimination does not transfer from vertical to horizontal orientations. Ball and Sekuler (1982), whose paradigm is used in our own experiments, have found that learning in motion discrimination in one pair of directions does not transfer to another pair of directions 90° or further away. Karni and Sagi (1991) have found that learning in a visual search task does not transfer when the background line elements rotate 90°. Finally, Poggio, Fahle and Edelman (1992) have found that learn-

* Corresponding author. Fax: +1-973-353-1917.

E-mail addresses: liu@psychology.rutgers.edu (Z. Liu), daphna@cs.huji.ac.il (D. Weinshall)

PII: S0042-6989(99)00140-6

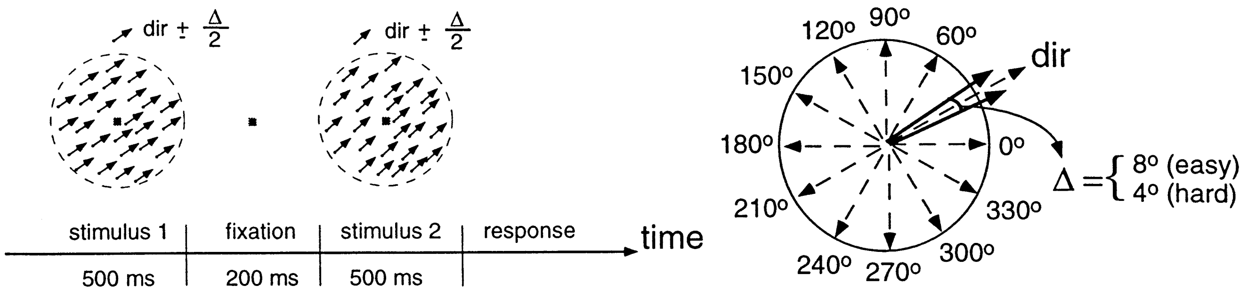


Fig. 1. Schematic illustration of one trial of motion discrimination. (Left) each stimulus consisted of a random dot pattern moving inside a circular aperture. The direction of each of the two stimuli was randomly chosen from two candidate directions: dir $\pm$ $\Delta/2$. The subject judged whether the two stimuli moved in the same or different directions. Feedback was provided. (Right) the fixed direction dir was chosen from 12 primary directions.

ing in Vernier acuity discrimination does not transfer from vertical to horizontal orientation and vice versa.

In our first experiment (Section 2.2), we use the same conventional paradigm, with which learning has been previously shown not to transfer. We, however, keep tracking the subjects' performance in the second training direction through time, and find that although the initial performance in the second direction is poor, learning nevertheless becomes faster. Thus, some knowledge learned in the first direction is used in the second. In other words, learning indeed generalizes in a task that does not show immediate transfer.

Some recent results showed that the aforementioned lack of transfer, or absence of immediate generalization, occurs only if the task is 'difficult'. Learning otherwise transfers when the task is 'easy.' For example, Liu (1995) has found, in motion discrimination, that a subject trained to discriminate 41 from 49° can later readily discriminate 131 from 139°. Ahissar and Hochstein (1997) have found, in a visual search task with line elements, that when the orientational difference between the target and distractor elements is enlarged, or when the target is restricted to only two possible locations, learning transfers when all elements are rotated 90°. In addition, Liu and Vaina (1995) have employed a simultaneous learning paradigm, in which learning trials in two pairs of motion directions are interleaved with unequal proportions. They found that the learning rate in the less frequent pair is greater than in the more frequent one, and suggested that learning transfers between the two interleaved tasks.

In our second experiment (Section 2.3) we replicate one such result as described in Liu (1995). One criticism of such results is that possibly there is no perceptual learning left in these 'easy' tasks, and that what is actually learned is some general knowledge about the task, whose transfer is to be expected.

A few recent experiments varied the perceptual learning paradigm yet in a different way, by manipulating task difficulty rather than attribute[1]. The exposure to an easy stimulus is found to facilitate learning of a more difficult stimulus with the same attribute. In a visual search task (Ahissar & Hochstein, 1997), long presentation of the stimulus in a single trial enables subjects to perform the task above chance when the stimulus presentation becomes short. Without this long presentation, subjects' performance remains at chance. This effect was termed *Eureka*. In a shape discrimination task (Rubin, Nakayama & Shapley, 1997), a session of training with easy stimuli enables subjects to learn the task with the difficult stimuli, which are difficult to learn otherwise. This effect was termed insight.

Our third experiment (Section 2.4) uses a hybrid design that manipulates both attribute (motion direction) and task difficulty. We found a new generalization mode, which we termed rooting. Both extensive training with one stimulus and a short exposure to another were needed for such generalization.

Thus, in contrast to previous results of specific learning, we show in three experiments that learning in motion discrimination often generalizes. The mode of generalization varies: (1) When the task is difficult, it is motion direction specific in the traditional sense, but learning in a new direction accelerates. (2) When the task is easy, learning generalizes to all directions after training in only one. (3) When subjects learn an easy task in one direction, and then practice briefly the same easy task in a new direction, learning generalizes to a difficult task in this new direction. While (2) is consistent with previous findings (Liu, 1995; Ahissar & Hochstein, 1997), (1) and (3) demonstrate that generalization is the rule, not an exception limited only to 'easy' stimuli.

The specificity of learning was used to support the hypothesis that perceptual learning embodies neuronal modifications in the brain's early stimulus-specific cortical areas (e.g. visual area MT) (Ramachandran, 1976; Gilbert, 1994; Karni, 1996). Our results show that

[1] We note that task difficulty, which is defined in motion discrimination as the angular difference between two possible motion directions, can also be considered as an attribute. In this paper, however, we distinguish task difficulty from other physical task attributes such as motion direction or line orientation.

generalization is common in perceptual learning, hence this hypothesis needs to be revised.

In Section 3, we adopt a signal detection framework to analyze our perceptual learning results. We describe the model's assumptions, including the limited computational capacity, and show simulation results that concur with human learning. Our approach cannot model the biological architecture, but it can demonstrate that a 'generic' discrimination system exhibits similar behavior to human subjects. From this similarity we conclude that the current psychophysical results, though very informative about human perceptual learning, cannot be used to identify the exact mechanism of perceptual learning in the brain.

2. Perceptual learning experiments

Below we describe three psychophysical experiments that demonstrate three different modes of generalization in perceptual learning. Experiment 1 in Section 2.2 shows acceleration of learning when the task is difficult. Experiment 2 in Section 2.3 shows transfer to all directions from one trained direction when the task is easy. Experiment 3 in Section 2.4 shows transfer after subjects have learned an easy task in one direction, and practiced briefly the same easy task in a new direction.

2.1. Methods

We start by describing the experimental task that is common to all three experiments.

2.1.1. Stimuli and apparatus

The motion discrimination task is described in Fig. 1. In each trial, the subject is presented with two consecutive stimuli, each moving in one direction that is randomly chosen from the following two: dir $+ (\Delta/2)$ and dir $- (\Delta/2)$ (dir is fixed, denoting the average of the two possible directions). Specifically, the stimuli were presented on a computer monitor (Silicon Graphics Indigo 2 Extreme, 60 Hz) with a resolution of 1280×1024 pixels. Each stimulus consisted of 400 dots that were randomly distributed inside a circular aperture, whose diameter was 342 pixels, or 8° in visual angle at a viewing distance of 60 cm. Each dot was 3×3 pixels that was approximately 4 min in visual angle. The dots were white (3.9 cd/m^2) on dark background (1.0 cd/m^2, contrast 59%). All the dots in one stimulus moved in the same direction, with a constant speed of 10 °/s. Subjects were asked to fixate on a red square (0.1 cd/m^2, 0.58, 0.35) 11 pixels (16 min) wide.

We manipulated two variables. The first was dir, the average of the two possible motion directions. It was randomly selected from 12 primary directions around the clock, corresponding to 30°/step starting from 0° (Fig. 1). The second variable was Δ, the angular difference between the two possible motion directions, determining the difficulty of the task. We used $\Delta = 4°$ for the difficult condition, $\Delta = 6°$ for the intermediate condition, and $\Delta = 8°$ for the easy one. The only reason we used $\Delta = 4°$, as opposed to the 3° used in Ball and Sekuler (1987) on an oscilloscope, was that our monitor's resolution is lower. We found in our pilot study that many subjects were unable to learn with $\Delta = 3°$. $\Delta = 8°$ was chosen as the easy condition because most subjects found it relatively easy to learn, yet still needed training to achieve good performance. We observed large variation between subjects in the amount of training needed to learn the task: for some subjects $\Delta = 6°$ was still 'easy,' for others it was already 'difficult.'

2.1.2. General procedure

In each trial, the first motion stimulus was presented for 500 ms. This was followed by a 200 ms interval, during which only the fixation mark was present. Then the second stimulus was presented for 500 ms. The subject fixated on the fixation mark, and decided whether the motion directions of the two stimuli were the same or different by pressing a mouse button. Feedback was always provided: a correct response was followed by an electronic beep, and silence otherwise. The next trial automatically started thereafter. Unless otherwise specified, each experimental session consisted of 700 trials that lasted for about 25 min. The subject viewed the stimuli binocularly. All the experiments were conducted in a dark room.

2.2. Experiment 1: a difficult task

Each subject was trained extensively in one primary direction, and then in another. Three subjects were trained with $\Delta = 4°$, two with $\Delta = 6°$. We compared the learning rate between the first and second primary motion directions.

2.2.1. Procedure ($\Delta = 4°$)

Two naive subjects (DJ and ZJX) and author ZL participated in the experiment. Subjects ZL and DJ needed 20 sessions (700 trials per session) of training in the first direction, and only nine sessions in the second direction, which was 90° away from the first (Fig. 2). Subject ZJX, who had no previous experience with any psychophysical experiments, needed seven sessions in the first direction, and only four in the second. Training stopped in the first direction when a subject's performance reached a plateau, and in the second direction when the performance matched the asymptotic level of the first direction. At the very beginning of the experiment, to reduce any effect of task familiarization, the subjects practiced the task in the first direction until they were sufficiently familiar with it, which amounted to about 20 trials per subject.

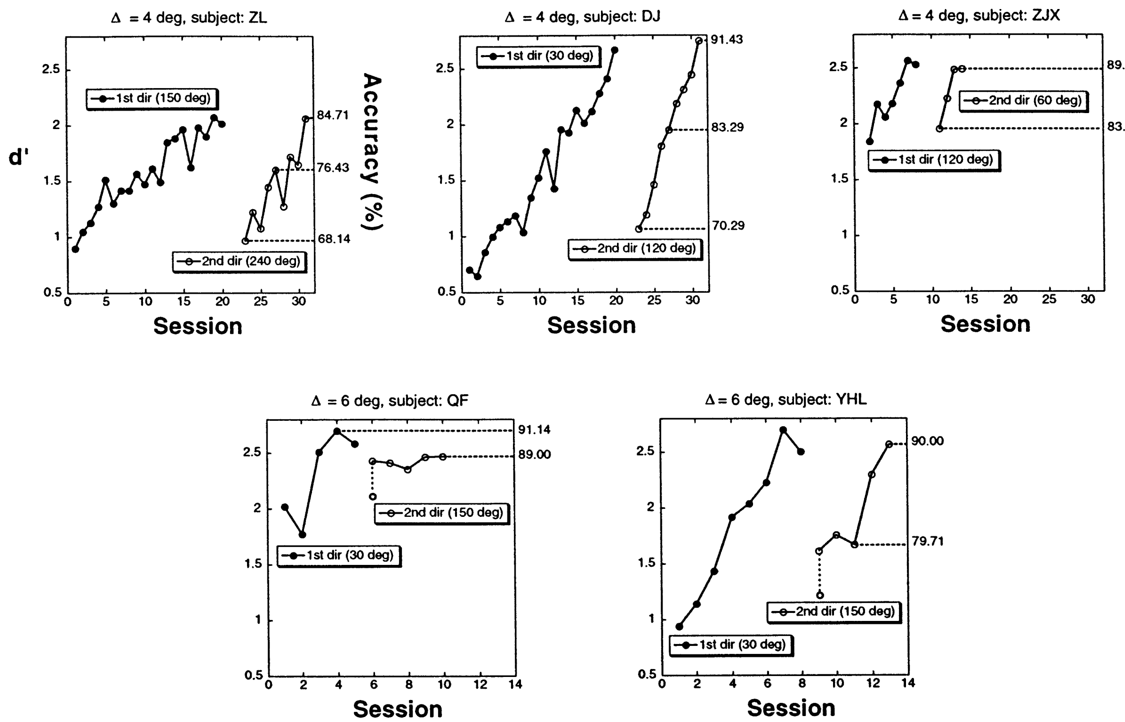


Fig. 2. (Top) Discrimination sensitivity d' as a function of training sessions. For comparison, a few accuracy scores are plotted on the right: (hits + correct rejections)/(number of trials). Subjects DJ and ZL needed 20 training sessions in the first direction, and nine in the second. Subject ZJX needed seven in the first, and four in the second. The rate of learning (the amount of improvement per session) nearly doubled in the second direction. (Bottom) same as above but with $\Delta = 6°$. In addition, each subject's baseline performance in the two directions was measured first. The baseline performance of the second direction is shown with a circle, connected by a dashed line to the second learning curve. Subject QF nearly transferred completely from the first to the second direction, whereas subject YHL's initial performance dropped half-way.

2.2.2. Results ($\Delta = 4°$)

The subjects' initial performance in the two directions was comparable, replicating the classical result of stimulus specific learning (no direct transfer). We computed linear regression for each learning curve and for each subject, to obtain the slope or learning rate $\Delta d'/$ session. A within subjects t-test yielded a significant difference ($t(2) = 6.40$, $P < 0.01$, one-tailed)[2]. The learning rate nearly doubled in the second direction (ratio, 1.96 ± 0.15). This indicates that although perceptual learning did not directly transfer in this difficult task, it did nevertheless generalize to the new direction. The generalization was manifested as nearly 100% increase in the rate of learning in the second direction.

2.2.3. Procedure ($\Delta = 6°$)

Since no direct transfer was found in the difficult condition ($\Delta = 4°$), while complete transfer has been found in the easy condition ($\Delta = 8°$) (Liu, 1995), we tested two additional naive subjects with an intermediate condition ($\Delta = 6°$). The experiment was identical to the above except for the following. At the beginning, the subjects' baseline performance was measured in both primary directions. This was done by interleaving blocks of trials from the two directions to avoid any ordering effect, with 50 trials in each block, and 700 trials total per subject.

2.2.4. Results ($\Delta = 6°$)

As expected, the subjects' performance was intermediate between the difficult and easy conditions (bottom of Fig. 2). Of special interest are individual differences: subject QF showed nearly complete transfer from the first to the second direction, as is typical of an 'easy' task. In contrast, subject YHL showed partial transfer. This demonstrates the relative nature of 'easy' and 'difficult' tasks. It also demonstrates that the two observed modes of generalization, direct transfer versus acceleration of learning, may be two extremes of a continuum.

2.3. Experiment 2: an easy task

In this experiment, we replicated results originally reported in Liu (1995) using the easy condition $\Delta = 8°$. We demonstrate that training in one direction can readily transfer to untrained directions.

[2] As there are only three subjects, it is impossible for the Wilcoxon matched-pair test to yield a significant result, even under the most favorable situation as is here.

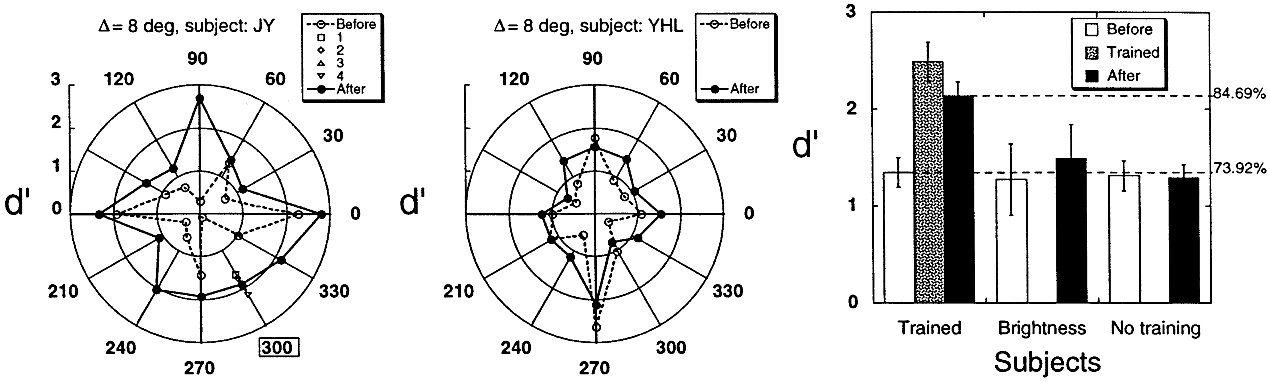


Fig. 3. (Left) Discrimination sensitivity *d'* of subject JY who was trained in the primary direction 300°. (Middle) *d'* of control subject YHL who had no training in between the two measurements around the clock. (Right) Average *d'* (and standard error) for all subjects before and after training. *Trained*: results for the four trained subjects. Note the substantial improvement between the two measurements. For these subjects, the *d'* measured after training is shown separately for the trained direction (middle column) and the remaining directions (right column). *Control I*: results for the three subjects who discriminated brightness during the training. *Control II*: results for the seven subjects who skipped training.

2.3.1. Procedure

We first measured the subjects' baseline performance in the 12 primary directions, 64 trials each (768 trials total). To avoid any ordering effect, the different directions were randomly interleaved. As a result the primary direction changed randomly from trial to trial in the baseline measurement, making the baseline task more difficult than the training task (where the primary direction is fixed). After the baseline measurement, each subject was trained in one oblique primary direction for four sessions, 700 trials each. The primary direction during training was randomly chosen for each subject and counter-balanced across subjects. Finally we measured the subjects' performance in all directions again, the same as in the baseline measurement.

Three naive subjects and author DW participated in the experiment. Our hypothesis was that training in one direction enables subjects to improve in all directions. To check whether any observed improvement is indeed due to training in one direction, we conducted a control experiment with seven subjects who did not undertake any training. The experiment was otherwise identical[3].

To further check whether such putative improvement was due to training in motion discrimination per se or only due to exposure to the motion stimuli, we conducted another control experiment with three additional naive subjects. Here, the brightness of the dots in a single trial could change from one stimulus to the next. The subjects' task was to decide whether the brightness of the dots was the *same* or *different* between the two stimuli, and feedback was provided (background, 1.0 cd/m^2; light, 3.9 cd/m^2, 59% contrast; dark, 2.5 cd/m^2, 43% contrast). We predicted that no transfer would occur if learning is specific to training in motion direction discrimination.

[3] Subject YHL also participated in Experiment 1, in the intermediate condition $\Delta = 6°$. Her data in the present experiment was collected first, however. As a control subject, YHL was tested equally often in all the primary directions. Therefore, we do not expect that YHL's performance in Experiment 1 was biased in any particular motion direction.

2.3.2. Results

Using the Wilcoxon rank-sum test, we found that the four trained subjects improved in nontrained directions significantly more than the seven control subjects ($\Delta d' = 0.79$ versus -0.02; $W_s(4, 7) = 10$, $P < 0.005$, one-tailed). The improvement of a trained subject was computed in all but the trained direction, and that of a control subject was computed in all directions. This result suggests that training with an easy task in one direction immediately improves performance in other directions. Hence the learned skill generalizes across motion directions (Fig. 3).

Next we compared the two control groups: the three subjects who did the brightness discrimination versus the seven who skipped training. It turned out that the improvement in motion direction discrimination of the first group was only marginally better than the second group ($\Delta d' = 0.21$ versus -0.02; $W_s(3, 7) = 11$, when $W_s(3, 7) = 10$, $P = 0.10$, one-tailed). This suggests that familiarization with the stimulus hardly helps subjects improve their discrimination. When the first three control subjects were compared with the four trained subjects, the difference was statistically significant ($\Delta d' = 0.21$ versus 0.79; $W_s(3, 4) = 6$, $P < 0.05$). This suggests therefore that exposure to the motion stimuli per se without direction discrimination is insufficient for direct transfer.

2.4. Experiment 3: generalization from easy to difficult stimuli

In previous studies, the subjects' exposure to an easy stimulus, termed *insight* or *Eureka*, has been found to facilitate learning of a more difficult task. Here we study the effect of such exposure in the context of extensive practice in a different direction. We have observed a new effect, which we term *rooting*.

2.4.1. Hypothesis

Experiment 3 was designed to determine whether the underlying mechanism of transfer is very rapid learning or 'real' transfer. If it is the latter, we expect that the subject knows the task before being tested with the new stimuli. If it is the former, then we do not expect that the subject knows the task beforehand but has to learn it very rapidly during the test. How can we distinguish these two possibilities? We cannot do so by direct measurement because very rapid learning, by definition, is faster than our measurements can detect. We therefore employed the rooting paradigm, with the following rationale.

Given three experimental conditions $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{C}$, if we know in advance that mastering $\mathscr{A}$ transfers to $\mathscr{B}$ and that $\mathscr{B}$ transfers to $\mathscr{C}$, then only with 'real' transfer does mastering $\mathscr{A}$ transfer directly to $\mathscr{C}$, whereas with very rapid learning mastering $\mathscr{A}$ does *not* immediately transfer to $\mathscr{C}$. In our rooting paradigm we let:

- $\mathscr{A}$ be the easy task in the direction with extensive training;
- $\mathscr{B}$ be the easy task in a different direction with brief training. From Experiment 2 we know that mastering $\mathscr{A}$ transfers to $\mathscr{B}$;
- $\mathscr{C}$ be the difficult task in the direction of $\mathscr{B}$. From Rubin et al. (1997) we deduce that $\mathscr{B}$ transfers to $\mathscr{C}$, since both are in the same direction.

We tested the hypothesis with three motion directions. The first is called the trained direction, in which a subject practices extensively with the easy task ($\mathscr{A}$). After this extensive practice, the subject is tested in two new directions. In one, which we call the *rooted* direction, the subject practices briefly with the easy task ($\mathscr{B}$), and is then tested with the difficult task ($\mathscr{C}$). In the other, which we call the *novel* direction, the subject is tested with the difficult task ($\mathscr{C}$) directly without practicing the easy task ($\mathscr{B}$). We make the following predictions:

If transfer is rapid learning, and after $\mathscr{A}$ is mastered:

- In the *novel* direction, there is no exposure to $\mathscr{B}$, thus the transfer $\mathscr{A} \rightarrow \mathscr{B}$ does not materialize. Hence, when the $\mathscr{C}$ stimuli are presented, there is no transfer $\mathscr{B} \rightarrow \mathscr{C}$.
- In the *rooted* direction, since the $\mathscr{B}$ stimuli are presented, the transfer $\mathscr{A} \rightarrow \mathscr{B}$ materializes. Hence, when the $\mathscr{C}$ stimuli are presented, there is transfer $\mathscr{B} \rightarrow \mathscr{C}$.

Thus, if transfer is rapid learning, we predict that the performance with the difficult stimuli is good in the *rooted* direction and poor in the *novel* direction.

On the other hand, if transfer is real, then via transfer subjects already know the stimuli before being tested. Therefore after subjects have mastered $\mathscr{A}$ in the *trained* direction, $\mathscr{A}$ transfers to $\mathscr{B}$ in both the *rooted* and *novel* directions. We now predict that in both the *rooted* and *novel* conditions the discrimination of $\mathscr{C}$ should be good (similar to the discrimination of the *trained* direction).

2.4.2. Procedure

The subjects were trained to master the easy task ($\Delta = 8°$) in the *trained* direction (nine sessions, 700 trials each) (Fig. 4). They then practiced the easy task in a second direction (the *rooted* direction) for 100 trials. Finally, they were tested with the difficult task ($\Delta = 4°$) in the *trained*, the *rooted*, and a third direction (the *novel* direction) to which they had no previous exposure (200 trials in each direction). The order of the tests in the three directions was counter-balanced across subjects. Three control subjects went through the same procedure except that they skipped the extensive training in the trained direction. In other words, they did not master the easy task.

For the trained subjects, we predict that there should be a difference between the *rooted* and *novel* directions, which is due to the small number of trials (100) with the easy task in the *rooted* direction. The control experiment tests whether or not the small number of easy trials in itself is responsible for this difference. If the control subjects do not show a difference between the *rooted* and *novel* directions, then the extensive practice in the trained direction *and* the small number of easy trials in the *rooted* direction both account for this effect. Note that we do not need a separate condition, where subjects skip the brief training, to show that extensive training alone cannot account for this effect, since this condition is equivalent to the novel condition.

2.4.3. Results

For the trained subjects, their performance in the easy task was similar in the *trained* and *rooted* directions, as predicted by Experiment 2, which showed that mastering an easy task transfers directly to untrained directions.

In the difficult task, subjects' performance in the *trained* and *rooted* directions was not different ($d' = 0.49$ versus 0.43, $t(2) = 0.68$, $P = 0.28$). But they were significantly better in the *rooted* than in the *novel* directions ($d' = 1.32$ versus 0.86, $t(2) = 2.84$, $P < 0.05$). In contrast, the performance of the control subjects in the *rooted* and *novel* directions was *not* significantly different ($d' = 0.50$ versus 0.63; $t(2) < 0$). In addition, in the *rooted* direction, the trained subjects were significantly better than the control subjects (Wilcoxon rank-sum test: $W_s(3, 3) = 6$, $P < 0.05$, one tailed, Fig. 4). Thus a mastered easy task can generalize to a more

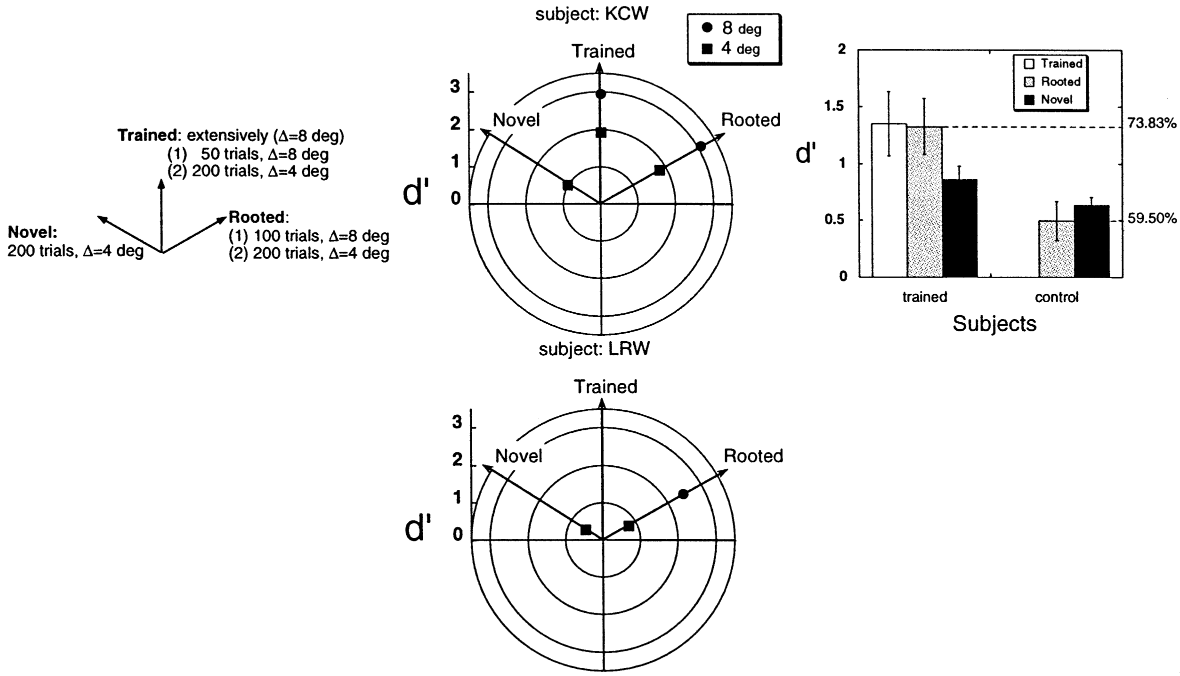


Fig. 4. Experiment 3. (Left) Schematic illustration of the experiment. After extensive training in one (trained) direction with an easy task $\Delta = 8°$, the subject was tested in the same direction with both the easy task (50 trials) and the difficult task ($\Delta = 4°$, 200 trials). In a second (*rooted*) direction, the subject was first tested with the easy task (100 trials) and then with the difficult task (200 trials). In a third (*novel*) direction, the subject was only tested with the difficult task (200 trials). The three directions were randomly chosen for each subject, with the constraint that the *rooted* and *novel* directions are symmetric about and 90° away from the *trained* direction. The order of the three difficult tests were counter balanced for the trained subjects. Three control subjects, who skipped the initial extensive training, repeated the same tests in the same counter balanced order, in the rooted and novel directions only. (Middle) Discrimination sensitivity d' for trained subject KCW and control subject LRW. (Right) Average d' (and standard error) for the trained (left) and control (right) subjects in the difficult task ($\Delta = 4°$).

difficult task in a new direction only after subjects have practiced, albeit briefly, the easy task in the new direction also. We term the brief practice with the easy task in the new direction rooting. Unlike *insight* and *Eureka*, rooting on its own cannot facilitate learning of the more difficult task.

2.4.4. Conclusions

Based on the discussion above our results indicate that rapid learning accounts for the observed transfer, since subjects' performance differs between the *rooted* and *novel* directions.

3. A computational model

We adopt a signal detection framework to analyze human perceptual learning. Our model accounts for the results in this paper by employing the constraint of limited computational capacity. Although some strong assumptions are made, e.g. that the measurements are taken from normal distributions, we ask the reader to keep our goal in mind: it is not to model the biological architecture but rather to show that a 'generic' discrimination system can demonstrate similar behavior to perceptual learning. This weak statement already justifies our argument that our psychophysical results, although informative about human perceptual learning, cannot be used to identify the exact mechanism of perceptual learning in the brain.

The model's specifics are as follows. Each experiment consists of two consecutive sessions of a discrimination task, which, for ease of reference to the psychophysical results in the previous section, we call the two primary motion directions. Each of the two sessions includes a large number of trials. In each trial two stimuli are presented, and the model decides whether their motion directions are the same or different. Our model's assumptions are as follows.

1. In each trial, each of the two stimuli is represented by a set of measurements that encode all aspects of the stimulus (e.g. the output of a localized direction detector). The measurements are encoded as a vector. Whether the two stimuli are the same or not is determined by the difference of the two vectors.
2. Each component of a measurement vector is characterized by its sensitivity to the discrimination task, i.e. how well the two stimuli can be discriminated by this component alone. The measurements are divided into a few discrete sets (e.g. direction selective versus speed selective). Pertinent to the task, each set of measurements is either *informative* — mea-

surements that have significant sensitivity, or *non-informative* — those that do not. Nevertheless, an *informative* component may vary greatly in its sensitivity value, from zero to very high. When many measurements have high sensitivity, the task is easy. Otherwise, when only a few do, the task is difficult.

We assume that the sensitivity of an individual component changes from one motion direction to the next, but the set of *informative* measurements remains unchanged. For example, in our task localized directional signals are *informative*, though the sensitivity of each component varies dependent on specific motion directions. On the other hand, local speed signals are never *informative*.

3. Due to the limited computational capacity, the model can simultaneously process only a small subset of the measurements from the input vector. The decision in a single trial is therefore based on the magnitude of this sub-vector, which is sampled from the measurement vector. The sampling is as follows.

   In each trial the model ranks the processed subset of measurements according to their sensitivity to the task. After a sufficient number of trials (enough to estimate the sensitivity of the processed subset), the model identifies the least sensitive measurement and replaces it with a new random component from the input vector. In effect, the model is searching for a subset of the input vector that gives rise to the maximal discrimination sensitivity. Consequently, the performance of the model is gradually improving, which corresponds to learning from trial to trial in each motion direction.
4. After learning in the first direction, the model identifies the *informative* and *non-informative* sets of measurements. A set is identified as *informative* if some of its sampled measurements had significant (though possibly low) discrimination sensitivity in the first training direction. In the next training direction, only the set of *informative* measurements is searched. The search therefore becomes more efficient, and hence the acceleration of learning. This accounts for the learning between training directions. Thus the learning rate is predicted *not* to increase with exposure only. In other words, subjects need to discriminate (and not just to be exposed to) the stimuli for effective inter-directional generalization.

In the following we will further specify these assumptions, and show simulations of different learning phenomena.

3.1. Learning with limited computational capacity

3.1.1. Notations

We assume that each stimulus generates an input that is a vector of N measurements: $\{I_i\}_{i=1}^N$. We also assume that the discrimination is based on the difference between the two stimuli in a trial: $\mathbf{x} = \{x_i\}_{i=1}^N$, $x_i = \Delta I_i$. The discrimination task amounts to deciding whether $\mathbf{x}$ is generated by noise — the null vector $\emptyset$, or by some signal—the vector $\mathscr{S}$.

At time t a measurement vector x^t is obtained, which we denote as x^{st} if it is the signal $\mathscr{S}$, and x^{nt}, otherwise. Assume that each component in x^t is a normal random variable:

$$x^{nt} = \{x_i^{nt}\}_{i=1}^N : x_i^{nt} \sim N(0, \sigma_i),$$
$$x^{st} = \{x_i^{st}\}_{i=1}^N : x_i^{st} \sim N(\mu_i, \sigma_i). \quad (1)$$

We measure the sensitivity d' of each component. Since both the signal and noise are normal random variables, the sensitivity of the i-th measurement in the discrimination task is $d'_i = |\mu_i|/\sigma_i$. Assuming further that the measurements are independent of each other and of time, the combined sensitivity of M measurements is

$$d' = \sqrt{\sum_{i=1}^{M} \left(\frac{\mu_i}{\sigma_i}\right)^2}. \quad (2)$$

Traditionally and without feedback, the sensitivity d' determines (or bounds) asymptotic performance of a discrimination system. In the next section we discuss how to modify this framework to accommodate learning with limited computational capacity. The effects of feedback are discussed in Appendix A.

3.1.2. Limited computational capacity: an assumption

We assume that the model can simultaneously process at most $M \ll N$ of the original N measurements. Since different components have different sensitivity, the discrimination depends on the combined sensitivity of the sampled M measurements. Learning in the first training direction, therefore, leads to selecting a 'good' subset of the measurements via searching. In the first training direction the entire measurement space is searched.

Having found the best M measurements in the first training direction, the model divides the measurements into two sets: those with above threshold sensitivity, and those without. The first set is identified as *informative*. Such ranking is used for the next training direction, when only the *informative* set is searched.

To illustrate the limited capacity assumption, consider the following example. Assume that the model has N measurements of two types: $N/2$ motion direction measurements and $N/2$ speed measurements. The model learns during the first training direction that all sampled speed measurements have null sensitivity, whereas the sampled directional measurements have varying (but usually significant) sensitivity (Fig. 5).

In the second training direction, the model receives N measurements whose sensitivity distribution differs from that in the first training direction, but still only

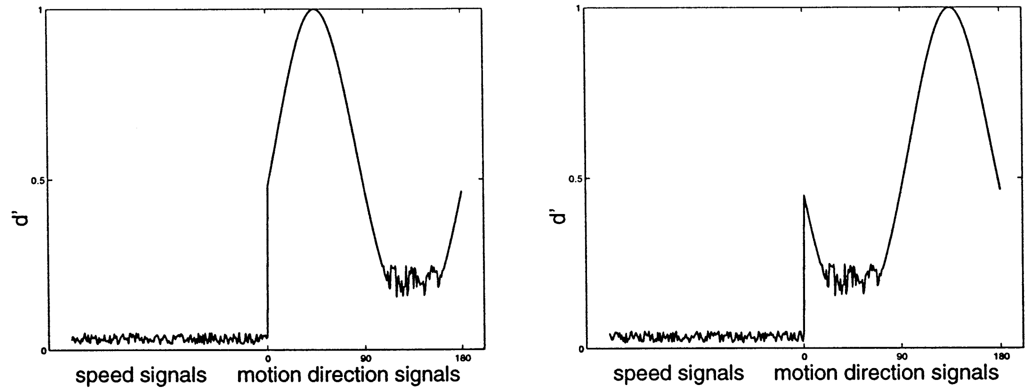


Fig. 5. Hypothetical sensitivity profile for a population of measurements of speed and motion direction. (Left) First training direction — only some of the motion direction measurements have significant sensitivity (say, d' above 0.5), with measurements around 45° having the highest d'. (Right) Second direction — only some motion direction measurements have significant sensitivity, with measurements around 135° having the highest d'.

directional measurements have significant sensitivity (Fig. 5). Being trained in the first direction, the model now only searches the measurements that were *informative* in the first direction, namely, the directional measurements. Now the asymptotic performance in the second direction remains unchanged because the most sensitive measurements are within the searched subset. The learning rate, however, doubles as the model searches a space half as large[4].

3.1.3. What determines task difficulty: an assumption

To account for the different modes of learning, we make the following assumption. When the task is difficult, only a small number of informative components have high d'. When the task is easy, many *informative* measurements have high d'. Therefore, when the task is easy, a set of M measurements that gives rise to the best performance is found relatively fast, especially after the *informative* measurements have been identified. In the extreme, when the task is very easy (e.g. all *informative* measurements have high sensitivity), learning is almost instantaneous and the outcome appears like transfer. On the other hand, when the task is difficult, it takes much longer to find the best M measurements. Hence learning is slow.

3.1.4. Simulation protocol

The detailed simulations of the model are as follows. In the first training direction, the model starts with a random set of M measurements. In each trial and with feedback, the mean and standard deviation of each measurement is estimated: μ_i^{st}, σ_i^{st} when the signal is present, and μ_i^{nt}, σ_i^{nt} otherwise. In the next trial, using the same subset of M measurements $\{x_i^{t+1}\}_{i=1}^{M}$, the model evaluates

$$\delta = \sum_{i=1}^{M} \left\{ \left(\frac{x_i^{t+1} - \mu_i^{st}}{\sigma_i^{st}} \right)^2 - \left(\frac{x_i^{t+1} - \mu_i^{nt}}{\sigma_i^{nt}} \right)^2 \right\}, \tag{3}$$

and classifies $\mathbf{x}$ as the signal if $\delta < 0$, and as noise otherwise. After feedback, μ_i^{st}, σ_i^{st} are updated if the stimuli are different (signal), and μ_i^{nt}, σ_i^{nt} are updated if the stimuli are the same (noise).

At time T, the least useful measurement is identified as *argval* of

$$\min_i d'_i, \quad d'_i = \frac{|\mu_i^{sT} - \mu_i^{nT}|}{(\sigma_i^{sT} + \sigma_i^{nT})/2} \tag{4}$$

(see Appendix). It is then randomly replaced by one of the remaining $N - M$ measurements. The learning and decision making proceed as above for another T iterations. This is repeated until the combined sensitivity of the M chosen measurements stabilizes. At the end, the decision is made based on the set of M measurements that have the highest sensitivity.

At the end of training in the first direction, based on the estimated d'_i, the sampled measurements are labeled as *informative* — those with d'_i larger than some threshold, and *non-informative* otherwise. All non-sampled measurements that are in the same set as the *informative* are labeled as such. In the second direction the learning proceeds as above, but only the *informative* measurements are searched.

3.1.5. Simulation results

In the simulation we used $N = 150$ measurements, with $M = 4$, and two sets. Half of the N measurements (the first set) had above threshold d'_i, where the threshold was set arbitrarily to 1. As an example, the measurements with high d'_i had a mean sensitivity of 20 (in the range 15–25). In the second training direction,

[4] We reiterate that the 50–50 division of directional and speed signals serves as a hypothetical example for illustration only. We are not suggesting that there are half directional and half speed signals in the visual system. We do not intend to quantify the acceleration of the learning rate either.

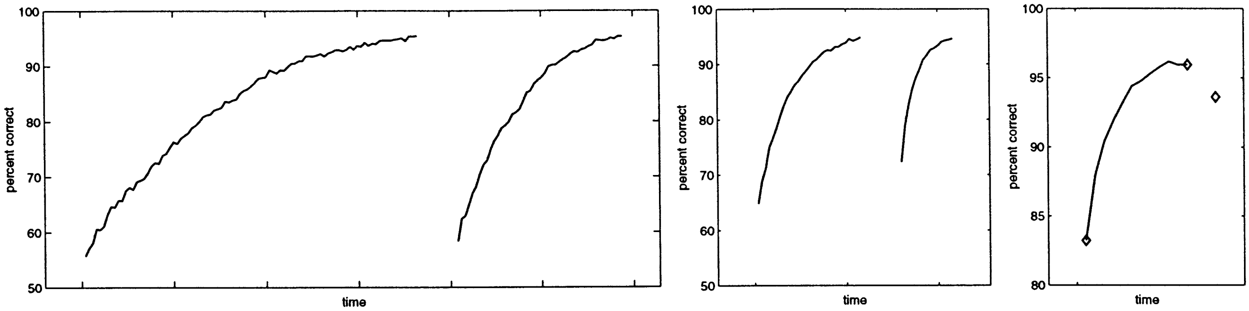


Fig. 6. Simulated performance (percent correct) as a function of time. (Left) Difficult task — the number of measurements with high d'_i is small (four out of 150). There is no transfer from the first to the second training direction, but the learning rate doubled. This graph is qualitatively similar to the results shown in the top row of Fig. 2. (Middle) Intermediate task — the number of measurements with high d'_i is larger (20 out of 150). There is partial transfer from the first to the second training direction, and the learning rate also increases. This graph is qualitatively similar to the results shown in the bottom row of Fig. 2. (Right) Easy task — the number of measurements with high d'_i is large (72 out of 150). There is almost complete transfer from the first to the second training direction.

the sensitivity of the measurements were randomly changed, while maintaining that only those in the first set have above threshold d'_i. By varying the number of direction sensitive measurements with high d'_i, and thus varying the difficulty of the task, we obtain the different modes of generalization (Fig. 6).

3.2. Incorporating Bayesian inference

So far we have focused on generalization between two directions that differ in the sensitivity of the components, but share the same level of difficulty (the same number of highly sensitive components). In other words, the sensitivity distribution over the population of measurements had different means, but approximately the same variance. We have observed the different modes of generalization described in Experiments 1 and 2 (Fig. 6).

The phenomena of *insight* and *Eureka*, and our rooting result in Experiment 3, deal with generalization from an easy to a difficult task with the same attribute. In other words, the sensitivity distribution in subsequent sessions peaks at approximately the same attribute, but has different variances (different number of components with above threshold sensitivities). The variance is large in an easy task, and is small in a difficult task.

To take advantage of the peak correspondence in subsequent sessions, we have incorporated Bayesian updating into the model. Specifically, the model uses the learned parameters (mean and standard deviation) of the distributions of $\mathcal{S}_i^T$ and $\emptyset_i^T$, which are computed in the first training direction, as Bayesian priors in the second training direction. As the most sensitive measurement remains the same throughout the two tasks, we obtain transfer from the first to the second direction because the estimation starts off from a better prior in the second direction. This is illustrated in Fig. 7a.

3.2.1. The model and Experiment 3

Our results in Experiment 3 (Section 2.4) indicate that rapid learning accounts for the observed transfer. Our model accordingly employs rapid learning as the mechanism underlying transfer.

The simulations of our model above showed transfer from one direction to another in the easy task. Using the notations of Section 2.4, $\mathcal{A}$ transfers to $\mathcal{B}$ (Fig. 6). By including Bayesian updating in our model, we also found transfer from the easy to difficult tasks in the same motion direction, i.e. $\mathcal{B}$ to $\mathcal{C}$ (Fig. 7a). By combining these two, our model can replicate the results of Experiment 3 as well.

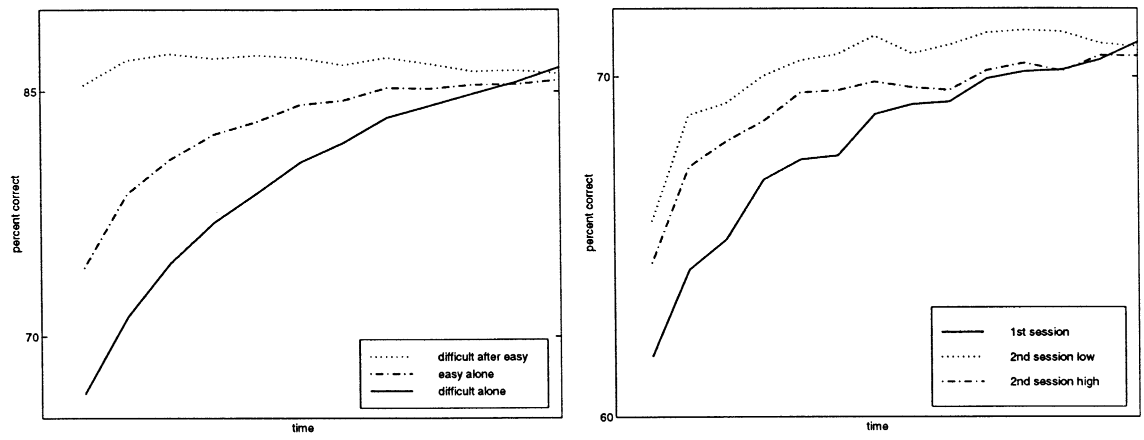
3.2.2. Is Bayesian inference sufficient?

It may appear that Bayesian inference suffices to explain Experiments 1 and 2 without the assumption of limited computational capacity. This is not the case, however, because in the first two experiments the parameters in the second direction are significantly different from the first. A prior based on the first direction at most improves the initial performance in the second direction (partial transfer) but does not improve the learning rate (Fig. 7b). Thus, Bayesian inference alone only predicts one aspect of the generalization — direct transfer, but not the acceleration of learning[5].

4. Summary and discussion

In contrast to previous results of stimulus specific learning (Ramachandran & Braddick, 1976; McKee & Westheimer, 1978; Vogels & Orban, 1985; O'Toole &

[5] We note that our model is partially Bayesian — we allow for individual measurement's sensitivity to change from one training direction to the next, but the informative set remains unchanged and can be estimated across training directions.



a) b)

Fig. 7. Simulation of a Bayesian decision model, whose learning rate changes from the first to the second motion direction. We show two cases here. (a) An easy task is learned first, and is followed by a difficult one. Both are in the same direction. Learning the difficult task alone takes a long time (lower plot). But when it follows the easy task — learning is instantaneous (upper plot). (b) The second task has the same level of difficulty as the first, but in a different direction: 2nd session low — the parameters of the measurements change only slightly from the first training direction to the second; 2nd session high — the change is larger. In both cases the performance in the second direction starts off better, and continues to improve as fast as it did in the first direction (at the same point on the horizontal axis). There is no acceleration of learning, however.

Kersten, 1992; Shiu & Pashler, 1992; Saarinen & Levi, 1995; Schoups, Vogels & Orban, 1995; Vaina, Sundareswaran & Harris, 1995; Matthews & Welch, 1997), we broadened the search for generalization beyond traditional transfer. We found that generalization is the rule, not an exception. Perceptual learning of motion discrimination generalizes in various ways: as an acceleration of learning (Experiment 1), as an immediate improvement in performance (Experiment 2), or as an immediate improvement after immediate transfer of a mastered skill (Experiment 3). In fact, the results in Experiment 1 suggest that the two observed modes of generalization, immediate improvement and acceleration of learning rate, may be two extremes of a continuum. Thus we demonstrate that perceptual learning is more similar to cognitive learning than previously thought, with both stimulus specificity and generalization as important ingredients.

To analyze the modes of generalization we adopted a signal detection framework, with the constraint of limited computational capacity. In this framework all the observed generalization modes are accounted for, with the same mechanism at work in both the ‘easy’ and ‘difficult’ tasks. We replicated the observed phenomena — transfer and the acceleration of learning rate — for respectively increasing levels of task difficulty. As in Experiment 3, our model does not predict transfer per se, but instead a dramatic increase in learning rate that is equivalent to transfer.

Our model is forced by the limited computational capacity to search the measurement space. Generalization — transfer and increased learning rate — occurs as search efficiency increases from one training direction to the next, when the search space decreases. Our model also predicts that the learning rate should improve only if the subject discriminates the relevant attribute dimension, in apparent accordance with the results of Experiment 1.

As for biological implications, the assumption that many directional selective cells respond distinctively (albeit slightly) to motion stimuli in any direction may not be entirely unreasonable. Note that the tuning curve of a directional selective cell in area MT is very wide, typically spanning 90° between half amplitudes, and is positive for more than 180°. Our assumption only requires that a cell responds to all directions, whereas the response may be very small and reliable only after hundreds of trials[6].

In contrast to the model proposed in Ahissar and Hochstein (1997), our model employs a single mechanism to account for all the existing modes of perceptual learning. At the same time, our model is qualitative and less specific, and thus does not make any concrete quantitative predictions. We would like to emphasize that this is not a handicap; our goal is to show, qualitatively, that the various generalization phenomena are not surprising, as they should occur in a generic

[6] There is one possible confound in the experiment that we cannot rule out, namely the role of eye movements. Although eye movements are apparently negligible in motion discrimination (see Ball & Sekuler, 1987), and although our subjects were instructed to fixate on the center mark of the stimulus, we cannot rule out the possibility of ‘covert eye movement’ (Georgopoulos, 1995). We thank D. Kersten for pointing this out.

discrimination system with limited computational capacity. We therefore argue that it may be too early to use existing perceptual learning results to identify cortical areas of perceptual learning, and the levels at which learning takes place (cf. Sowden, Davies, Rose & Kaye, 1996).

Acknowledgements

This work was done when ZL was a Visiting Scientist and DW was on sabbatical at the NEC Research Institute, Princeton, NJ. We thank Manfred Fahle, David W. Jacobs, Dan Kersten, Scott Kirkpatrick, Mike Langer, Pascal Mamassian, Nestor Matthews, Suzanne McKee, John Oliensis, and Bosco Tjan for helpful comments and discussions.

Appendix A

When feedback is available, we assume that the model can learn the parameters of the signal $\mathscr{S}$ — the distribution of the vector x^{st}, and the noise — the distribution of the vector x^{nt}. At time t the model compares the input measurements with the estimated signal and noise. For simplicity, we assume here that there is only one signal and one measurement, thus the index i can be dropped in the following discussion.

Let $\mathscr{S}^t$ denote the estimated signal — the mean of the observations $x^{st'}$ up to time t, and the estimated noise similarly $\emptyset^t$ — the mean of the observations $x^{nt'}$ up to time t. Thus $\mathscr{S}^t$ and $\emptyset^t$ are normal random variables, where

$$\emptyset^t = \langle x^{nt'} \rangle_{t'=0\ldots t} \sim N\left(0, \frac{\sigma}{\sqrt{t_n}}\right)$$
$$\mathscr{S}^t = \langle x^{st'} \rangle_{t'=0\ldots t} \sim N\left(\mu, \frac{\sigma}{\sqrt{t_s}}\right) \quad (5)$$

for t_s, t_n the number of occurrences of the signal and noise, respectively. In the first t presentations, $t_s + t_n = t$. (The above follows from the distribution of the mean and standard deviation of i.i.d. normal random variables.)

The discrimination depends on the difference between the input signal x^t, and the estimated mean signal $\mathscr{S}^t$ and noise $\emptyset^t$. The likelihood of miss (reporting 'same' when stimuli are different) and false alarm (reporting 'different' when same) is inversely proportional to l_{miss} and l_{fa} respectively, which are defined as follows:

$$l_{\text{miss}} = d'(\mathbf{x}^{st}, \emptyset^t) - d'(x^{st}, \mathscr{S}^t)$$
$$l_{\text{fa}} = d'(x^{nt}, \mathscr{S}^t) - d'(x^{nt}, \emptyset^t). \quad (6)$$

To estimate l_{miss} and l_{fa}, we compute d' for two normal random variables with different standard deviations, in a yes/no decision task with no bias (the likelihood ratio equals to 1). Let $\mathbf{x} \sim N(\mu_x, \sigma_x)$ and $\mathbf{y} \sim N(\mu_y, \sigma_y)$ denote two normal random variables, where w.1.o.g. $\mu_y > \mu_x$. A likelihood ratio of 1 is obtained at point u when

$$\frac{f_x(u)}{f_y(u)} = 1 \Rightarrow \frac{(u-\mu_x)^2}{2\sigma_x^2} = \frac{(u-\mu_y)^2}{2\sigma_y^2} \Rightarrow u = \mu_x \frac{\sigma_y}{\sigma_x+\sigma_y} = \mu_y \frac{\sigma_x}{\sigma_x+\sigma_y}. \quad (7)$$

We assume that u is the decision boundary, so that the answer is $\mathbf{y}$ for observed values larger than u and $\mathbf{x}$ otherwise. Then, we have

$$d' = \frac{u-\mu_x}{\sigma_x} + \frac{\mu_y - u}{\sigma_y} = 2\,\frac{\mu_y-\mu_x}{\sigma_x+\sigma_y} = \frac{|\Delta\mu|}{(\sigma_x+\sigma_y)/2}, \quad (8)$$

where $d' = \mathscr{Z}$(hit rate)—$\mathscr{Z}$(false alarm rate), and $\mathscr{Z}()$ is the inverse of the normal distribution function. We plug Eq. (8) into Eq. (6) and obtain:

$$d'(x^{nt}, \emptyset^t) = 0, \quad d'(x^{nt}, \mathscr{S}^t) = \frac{|\mu - 0|}{\left(\sigma + \frac{\sigma}{\sqrt{t}}\right)/2} \quad (9)$$

$$\Rightarrow l_{\text{fa}} = l_{\text{miss}} = \frac{|\mu|}{\left(\sigma + \frac{\sigma}{\sqrt{t}}\right)/2} = d'(x^{st}, x^{nt})\,\frac{2}{\left(1 + \frac{1}{\sqrt{t}}\right)}. \quad (10)$$

The increase in sensitivity with time t reflects the learning within a training direction, and the asymptotic performance increases with the given $d'(x^{st}, x^{nt})$ of the measurement.

Note that in our modified model with feedback, the sensitivity starts as $d'(x^{st}, x^{nt})$ for $t = 1$, and increases asymptotically to $2d'(x^{st}, x^{nt})$. In other words, the asymptotic performance with feedback is twice as much as that without feedback, in qualitative agreement with experimental data (Ball & Sekuler, 1982). Since we are only interested in qualitative behavior and not quantitative predictions, we ignore this difference in our discussion, since the change of sensitivity l_{miss}, l_{fa} with time is negligible in comparison with the change due to the increase in search efficiency.

References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*, 401–406.

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science*, *218*, 697–698.

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, *27*(6), 953–965.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, *287*, 43–44.

Gilbert, C. D. (1994). Early perceptual learning. *Proceedings of the National Academy of Sciences USA*, *91*, 1195–1197.

Georgopoulos A. P. (1995). Current issues in directional motor control. *Trends in Neurosciences*, *18*.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Science USA*, *88*, 4966–4970.

Karni, A. (1996). The acquisition of perceptual and motor skills: a memory system in the adult human cortex. *Cognitive Brain Research*, *5*, 39–48.

Liu Z. (1995). *Learning a visual skill that generalizes*. Technical report, NEC Research Institute.

Liu, Z., & Vaina, L. M. (1995). Stimulus specific learning: a consequence of stimulus-specific experiments? *Perception*, *24*(supplement), 21.

Matthews, N., & Welch, L. (1997). Velocity-dependent improvements in single-dot direction discrimination. *Perception and Psychophysics*, *59*, 60–72.

McKee, S. P., & Westheimer, G. (1978). Improvement in vernier acuity with practice. *Perception and Psychophysics*, *24*, 258–262.

O'Toole, A. J., & Kersten, D. (1992). Learning to see random-dot stereograms. *Perception*, *21*, 227–243.

Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, *256*, 1018–1021.

Ramachandran, V. S. (1976). Learning-like phenomena in stereopsis. *Nature*, *262*, 382–384.

Ramachandran, V. S., & Braddick, O. (1976). Orientation-specific learning in stereopsis. *Perception*, *2*, 371–376.

Rubin, N., Nakayama, K., & Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. *Current Biology*, *7*, 461–467.

Saarinen, J., & Levi, D. M. (1995). Perceptual learning in vernier acuity: what is learned? *Vision Research*, *35*, 519–527.

Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularity. *Journal of Physiology*, *483*, 797–810.

Shiu, L., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics*, *52*(2), 582–588.

Sowden, P., Davies, I., Rose, D., & Kaye, M. (1996). Perceptual learning of stereoacuity. *Perception*, *25*, 1043–1052.

Vaina, L., Sundareswaran, V., & Harris, J. G. (1995). Learning to ignore: psychophysics and computational modeling of fast learning of direction in noisy motion stimuli. *Cognitive Brain Research*, *2*, 155–163.

Vogels, R., & Orban, G. A. (1985). The effect of practice on the oblique effect in line orientation judgments. *Vision Research*, *25*, 1679–1687.