Parameter learning but not structure learning: A Bayesian network model of constraints on early perceptual learning

Melchi M. Michel

Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, NY, USA

Robert A. Jacobs

Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, NY, USA

Visual scientists have shown that people are capable of perceptual learning in a large variety of circumstances. Are there constraints on such learning? We propose a new constraint on early perceptual learning, namely, that people are capable of parameter learning—they can modify their knowledge of the prior probabilities of scene variables or of the statistical relationships among scene and perceptual variables that are already considered to be potentially dependent—but they are not capable of structure learning—they cannot learn new relationships among variables that are not considered to be potentially dependent, even when placed in novel environments in which these variables are strongly related. These ideas are formalized using the notation of Bayesian networks. We report the results of five experiments that evaluate whether subjects can demonstrate cue acquisition, which means that they can learn that a sensory signal is a cue to a perceptual judgment. In Experiment 1, subjects were placed in a novel environment that resembled natural environments in the sense that it contained systematic relationships among scene and perceptual variables that are normally dependent. In this case, cue acquisition requires parameter learning and, as predicted, subjects succeeded in learning a new cue. In Experiments 2–5, subjects were placed in novel environments that did not resemble natural environments—they contained systematic relationships among scene and perceptual variables that are not normally dependent. Cue acquisition requires structure learning in these cases. Consistent with our hypothesis, subjects failed to learn new cues in Experiments 2–5. Overall, the results suggest that the mechanisms of early perceptual learning are biased such that people can only learn new contingencies between scene and sensory variables that are considered to be potentially dependent.

Keywords: perceptual learning, motion discrimination, Bayesian networks, cue integration, perceptual cue acquisition

Citation: Michel, M. M., & Jacobs, R. A. (2007). Parameter learning but not structure learning: A Bayesian network model of constraints on early perceptual learning. *Journal of Vision, 7*(1):4, 1–18, http://journalofvision.org/7/1/4/, doi:10.1167/7.1.4.

Introduction

Acquiring new information about the world requires contributions from both nature and nurture. These factors determine both what biological organisms can and cannot learn. Numerous studies have shown that organisms' learning processes are often biased or constrained. Perhaps the most famous demonstration that learning processes are biased comes from the work of Garcia and Koelling (1966) who showed that rats are predisposed to learn certain types of stimulus associations and not others. They interpreted their results in terms of a learning bias referred to as "belongingness"—organisms more easily learn associations among types of stimuli that are correlated or participate in cause-and-effect relationships in natural environments. A more recent demonstration that learning processes are biased comes from the work of Saffran (2002). She found that people more easily learned an artificial language when the statistical relationships among sound components were consistent with the dependencies that characterize the phrase structures of natural languages.

This article proposes a new constraint or bias on early, or low-level, perceptual learning.[1] We hypothesize that people's early perceptual processes can modify their knowledge of the prior probabilities of scene properties or their knowledge of the statistical relationships among scene and sensory variables that are already considered to be potentially dependent. However, they cannot learn new relationships among scene and sensory variables that are not considered to be potentially dependent, even when placed in novel environments in which these variables are strongly related. To illustrate this idea, consider the problem of perceptual cue acquisition. Wallach (1985) proposed a theory of cue acquisition that is representative of other theories in the literature (a closely related theory was originally proposed by Brunswick, 1956; readers interested in this topic should also see Haijiang, Saunders, Stone, & Backus, 2006). He hypothesized that in every perceptual domain (e.g., perception of motion direction), there is at least one primary source of information,

doi: 10.1167/7.1.4 Received April 18, 2006; published January 16, 2007 ISSN 1534-7362

useable innately and not modifiable by experience. Other perceptual cues are acquired later through correlation with the innate process. Using Wallach's theory, we consider constraints on the learning processes underlying cue acquisition.

One possibility is that these processes are general purpose, which means that they are equally sensitive to correlations between known cues and any signal. For example, let us suppose that retinal image slip is an innate cue to motion direction and let us consider an observer placed in a novel environment in which retinal image slip is perfectly correlated with a novel signal, such as the temperature of the observer's toes (e.g., leftward retinal slip is correlated with cold toes, and rightward retinal slip is correlated with hot toes). According to Wallach's theory, it ought to be the case that the observer learns that the temperature of his or her toes is a perceptual cue to motion direction. For example, the observer may learn that cold toes indicate leftward motion, whereas hot toes indicate rightward motion. Alternatively, it may be that the learning processes underlying cue acquisition are biased such that they are more sensitive to some correlations than to others. In particular, we conjecture that these processes cannot learn new relationships among scene and sensory variables that are not considered to be potentially dependent. It seems likely that an observer placed in the novel environment described above would not believe that motion direction and the temperature of his or her toes are potentially dependent variables, and thus, the observer's early perceptual system would fail to learn that the temperature of his or her toes is a cue to motion direction.

In the remainder of this article, we report the results of five experiments. These experiments evaluate our hypothesis regarding biases in early perceptual learning. They do so in the context of Wallach's theory of cue acquisition described above, namely, that new perceptual cues can be acquired by correlating an existing cue with a novel sensory signal. We then present a simple model, described in terms of Bayesian networks, that formalizes our hypothesis, accounts for our results, and is consistent with the existing literature on perceptual learning.

In Experiment 1, subjects were placed in a novel environment that resembled natural environments in the sense that it contained systematic relationships among scene and perceptual variables that normally share systematic relationships. Subjects were trained to perceive the motion direction of a field of moving dots when the visual cue to motion direction was correlated with a novel auditory signal. When an object moves in a natural environment, this event often gives rise to correlated visual and auditory signals. In other words, perceived auditory and visual motion signals are both dependent on the motion of objects in a scene and, thus, people regard visual or auditory signals as potentially dependent on the motion direction in a scene. We reasoned that subjects in our experiment should be able to estimate the motion direction of the moving dots based on the auditory and visual signals and then modify their knowledge of the relationship between motion direction and the initially unfamiliar auditory stimuli (i.e., to anticipate our discussion of Bayesian networks below, this would be an instance of parameter learning in which subjects modify their conditional distribution of the perceived auditory signal given the estimated motion direction). We predicted, therefore, that subjects would learn to use the hitherto unfamiliar and, thus, uninformative auditory stimulus as a cue to motion direction. As reported below, the experimental results are consistent with our prediction. Experiment 1 can be regarded as a control experiment in the sense that that it verified that our experimental procedures are adequate for inducing observers to learn a new perceptual cue in the manner suggested by Wallach (i.e., by correlating a signal that is not currently a cue with an existing cue).

In Experiments 2, 3, 4, and 5, subjects were placed in novel environments that did not resemble natural environments—they contained systematic relationships among scene and perceptual variables that do not normally share systematic relationships. In Experiments 2 and 3, the visual cue to motion direction was correlated with binocular disparity or brightness signals, respectively; the experimental procedures were otherwise identical to those of Experiment 1. In the natural world, neither brightness nor binocular disparity varies systematically with transverse object motion (i.e., motion in the frontoparallel plane). For example, it is not the case that brighter objects tend to move right whereas darker objects move left, nor is it the case that nearer objects tend to move right whereas distant objects move left. Consequently, observers should not consider motion direction and either brightness or binocular disparity as potentially dependent variables. In contrast to Experiment 1, the predictions of Wallach's hypothesis for Experiments 2 and 3 differ from those of our theory. Wallach's hypothesis suggests that correlating novel signals with existing cues should be sufficient to induce cue learning. In contrast, our hypothesis claims that observers can only learn relationships between variables that are considered to be potentially dependent. Because transverse motion direction and either brightness or binocular disparity are not considered to be potentially dependent, we predicted that subjects in Experiments 2 and 3 would fail to learn to use brightness or binocular disparity signals as cues to transverse motion direction (i.e., to anticipate the discussion of Bayesian networks below, we predicted that subjects would fail to show structure learning). The experimental results are consistent with this prediction.

Experiments 1, 2, and 3 attempted to teach subjects a new cue to transverse motion direction. To check that there is nothing idiosyncratic about this perceptual judgment, we used a different task in Experiments 4 and 5. Subjects were trained to perceive the light source direction when the shading cue to this direction was correlated with a visual disparity or auditory signal. Because neither binocular disparity nor auditory signals share systematic relationships with light source direction in the natural world, we predicted that subjects would fail to learn that

these signals were also cues to light source direction in our novel experimental environments. Again, the experimental results are consistent with this prediction.

Taken as a whole, the experimental results are consistent with the hypothesis that the learning processes underlying cue acquisition are biased by prior beliefs about potentially dependent variables such that cue acquisition is possible when a signal is correlated with a cue to a scene property and the signal is potentially dependent on that property. If the signal is not believed to be potentially dependent on the property, cue acquisition fails. In the General discussion section, we introduce a Bayesian network model formalizing this hypothesis.

Experiment 1: Auditory cue to motion direction

Subjects in Experiment 1 were trained to perceive the motion direction of a field of dots when the visual cue to motion direction was correlated with an auditory signal. The experiment examined whether subjects would learn that the auditory signal is also a cue to motion direction.

Because moving objects often give rise to both visual and auditory signals in natural environments (i.e., because sounds are created by physical motion), we expected that subjects would consider motion direction and an auditory signal to be potentially dependent and, thus, would learn that the auditory signal is also a cue.

Methods

Subjects

Subjects were eight students at the University of Rochester with normal or corrected-to-normal vision and normal hearing. All subjects were naive to the purposes of the study.

Stimuli

Visual stimuli were random-dot kinematograms (RDKs) presented for a duration of 1 s. The kinematograms consisted of 309 small antialiased white dots (each subtending approximately 0.65 min of visual angle) moving (at a rate of 1.4°/s) behind a simulated circular aperture (with a diameter of 5.72° of visual angle) against a black background. Half the dots in a display moved in the same direction, referred to as the stimulus direction, whereas each of the remaining dots moved in a direction sampled from a uniform distribution. Each dot had a lifetime of approximately 150 ms, after which a new replacement dot appeared in a random position within the aperture. These stimuli were presented on a standard 19-in. CRT with a resolution of 1,024 × 768 pixels and a refresh rate of 100 Hz and were viewed from a distance of 1.5 m. All experiments were conducted in a darkened room, with black paper obscuring the edges of the CRT.

Auditory stimuli consisted of 1 s of "notched" white noise played through a pair of headphones. We used auditory noise because we wanted to create ambiguous motion stimuli.[2] Two stimuli defining the endpoints of a continuum, denoted *A* and *B*, were each constructed by combining two narrow bands of noise (sampled at 22 kHz). Stimulus *A* had approximately equal amplitude in the ranges 4000–5000 and 8000–10000 Hz, whereas stimulus *B* had approximately equal amplitude in the ranges 1–2000 and 6000–7000 Hz. Intermediate stimuli were created by linearly combining stimuli *A* and *B*, where the linear coefficients formed a unit-length vector whose endpoint lied on a circle passing through the points (1,0) and (0,1) [e.g., the coefficients (1, 0) produced stimulus *A*, the coefficients (0, 1) produced stimulus *B*, and the coefficients $(1/\sqrt{2}, 1/\sqrt{2})$ produced a stimulus midway between *A* and *B*].[3] Auditory stimuli were normalized to have equal maximum amplitudes.

Procedure

The experiment used four tasks, referred to as the vision-only, audition-only, and vision–audition training tasks, and the vision–audition test task. The vision-only and audition-only tasks allowed us to characterize each subject's performances on visual and auditory discrimination tasks, respectively. The goal of the vision–audition training task was to expose subjects to an environment in which an auditory signal is correlated with a visual cue to motion direction. The goal of the vision–audition test task was to evaluate whether subjects learned that the auditory signal is also a cue to motion direction.

In each trial of the vision-only training task, four visual displays were presented: A fixation square was presented for 500 ms, followed by the first RDK for 1,000 ms, followed by a second fixation square for 400 ms, followed by the second RDK for 1,000 ms. The stimulus direction of the first RDK, referred to as the "standard" stimulus, was always 0° (vertical). The second RDK, referred to as the "comparison" stimulus, had a stimulus direction different from the standard. Subjects judged whether the dots in the comparison stimulus moved to the left (anticlockwise) or to the right (clockwise) of those in the standard (vertical) stimulus. They responded by pressing the appropriate key on the keyboard. At the end of every 10 trials, subjects were informed of the number of those trials on which they responded correctly. The ease or difficulty of the task was varied over trials by varying the stimulus direction of the comparison so that difficult trials contained smaller direction differences between the standard and comparison stimuli than did easy trials. This direction was determined using interleaved 2-up, 1-down and 4-up, 1-down staircases. Trials were run until there were at least 12 reversals of each staircase. A subject's approximate 71% correct and 84% correct thresholds were set to the average values over the last 10 reversals of the 2-up, 1-down and 4-up, 1-down staircases, respectively.

The audition-only training task was identical to the vision-only training task with the following exception. Instead of viewing RDK, subjects heard auditory stimuli. The standard was an auditory stimulus midway between stimuli A and B defined above, whereas the comparison was either nearer to A or nearer to B. Subjects judged whether the comparison was closer to A or B relative to the standard. Subjects were familiarized with A and B prior to performing the task.

Subjects also performed a vision–audition training task in which an auditory signal is correlated with a visual cue to motion direction. Before performing this task, we formed a relationship between visual and auditory stimuli by mapping subjects' visual thresholds onto their auditory thresholds. This was done using a log-linear function

$$\log(d_v) = m\log(d_a) + b, \quad (1)$$

where d_v and d_a are visual and auditory "directions," respectively, m is a slope parameter, and b is an intercept parameter. The log-linear function ensured that corresponding visual and auditory stimuli were (approximately) equally salient. The vision–audition training task was identical to the vision-only training task with the following exception. Instead of only viewing RDK, subjects both viewed RDK and heard the corresponding auditory stimuli. They were instructed to focus on the visual motion-direction discrimination task but were also told that the auditory stimulus might be helpful. Half the subjects were run in the "no-switch" condition, which means that the relationship between an auditory cue and a response key was the same on this task as it was on the audition-only task. The remaining subjects were run in the "switch" condition. (In other words, for half the subjects, the stimulus direction of auditory stimulus A was anticlockwise of vertical and the direction of B was clockwise of vertical, whereas this relationship was reversed for the remaining subjects.) This was done so that results on the vision–audition training and test tasks could not be attributed to an association between auditory stimuli and response keys learned when subjects performed the audition-only trials.

Vision–audition test trials were conducted to evaluate whether subjects learned that the auditory signal is correlated with the visual cue to motion direction and, thus, whether it, too, is a cue to motion direction. These test trials were similar to vision–audition training trials with the following differences. First, the presentation order of the standard and comparison was randomized. Subjects were instructed to judge whether the direction of the second stimulus was anticlockwise or clockwise relative to that of the first stimulus. Second, subjects never received feedback. Third, stimuli were selected according to the method of constant stimuli rather than according to a staircase. Importantly, standard stimuli were "cue-conflict" stimuli—the direction of the RDK was vertical, but the direction of the auditory stimulus was offset from vertical by either a value δ or $-\delta$, where δ was set to a subject's 84% correct threshold on the audition-only training trials. In contrast, the comparison stimulus was a "cue-consistent" stimulus. By comparing performances when the auditory signal in the standard had an offset of δ versus $-\delta$, we can evaluate whether this signal influenced subjects' judgments of motion direction.

Subjects performed the four tasks during two experimental sessions. In Session 1, they performed the vision-only and audition-only training tasks. Before performing these tasks, subjects performed a small number of practice trials in which they were given feedback on every trial. They also performed the vision–audition training task twice. In Session 2, they performed the vision–audition training task and then performed the vision–audition test task twice.

Results

Two subjects' results on the vision–audition test task are shown in Figure 1. The horizontal axis of each graph gives the direction of the comparison, whereas the vertical axis gives the probability that the subject judged the direction of the comparison as clockwise relative to that of the standard. The data points indicated by circles or crosses are for the trials in which the auditory signal in the standard was offset from vertical by the amount δ or $-\delta$, respectively. The dotted and solid lines are cumulative Normal distributions fit to these data points using a maximum-likelihood procedure from Wichmann and Hill (2001a).

To compare a subject's performances when the offset of the auditory signal in the standard was δ versus $-\delta$, we compared a subject's point of subjective equality (PSE) in each case. The PSE is defined as the direction of the comparison at which a subject is equally likely to judge this direction as being anticlockwise or clockwise relative to that of the standard. For example, consider subject BCY whose data are shown in the left graph of Figure 1. This subject's PSE is about $-3°$ in the $-\delta$ case and about $2°$ in the δ case, indicating a PSE shift of about $5°$. For each of the subjects whose data are illustrated in Figure 1, their PSE when the offset was $-\delta$ is significantly less than their PSE when the offset was δ (both subjects had significant PSE shifts using a significance level of $p < .05$, where the test of significance is based on a Monte Carlo procedure described by Wichmann & Hill, 2001b). Seven of the eight subjects run in the experiment had significant PSE shifts.

The graph in Figure 2 shows the combined data for all eight subjects, along with the maximum-likelihood psychometric fits for the pooled data. The average value of the offset δ across all subjects was equivalent to a $4.30°$ rotation in motion direction. Importantly for our purposes,

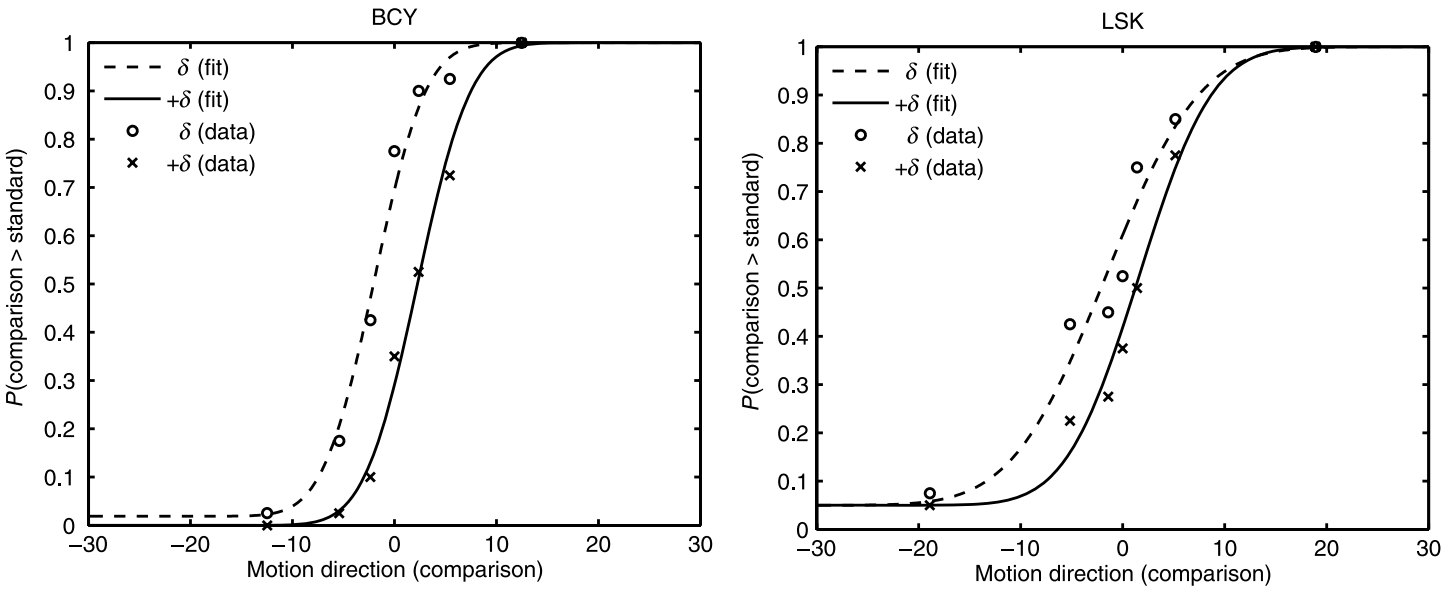

Figure 1. The data for two subjects from the vision–audition test trials. The horizontal axis of each graph gives the direction of the comparison, whereas the vertical axis gives the probability that the subject judged the direction of the comparison as clockwise relative to that of the standard. The data points indicated by circles or crosses are for the trials in which the auditory signal in the standard was offset from vertical by the amount δ or $-\delta$, respectively. The dotted and solid lines are cumulative Normal distributions fit to these data points using a maximum-likelihood procedure.

subjects showed a large shift in their PSE—the PSE shift for the combined data is 3.72° ($p < .001$). These data suggest that subjects based their judgments on information from both the visual and auditory signals. Had subjects used only the visual signal, we would have expected no shift in their PSEs. Conversely, if subjects had used only the auditory signal, then a PSE shift of 2δ (8.6° on average) would have been expected. The actual PSE shift (3.72° on average) was smaller, consistent with the idea that subjects combined information from the visual and auditory signals.

In summary, the results suggest that subjects acquired a new perceptual cue—they learned that the initially unfamiliar auditory signal was correlated with the visual cue to motion direction and, thus, it, too, is a cue to motion direction. Furthermore, the subjects used the new cue for the purposes of sensory integration—they combined information from the new auditory cue with information from the previously existing visual cue when judging motion direction.

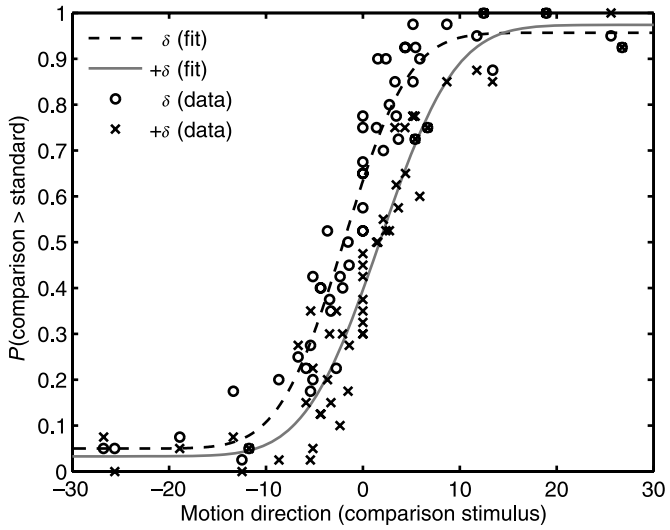

Figure 2. The data from the vision–audition test trials for all eight subjects combined.

Experiment 2: Disparity cue to motion direction

Subjects in this experiment were trained to perceive the direction of moving dots when the visual cue to motion direction was correlated with a binocular disparity signal. The experiment examined whether subjects would learn that the disparity signal is also a cue to motion direction. Because the transverse motion of objects in the natural world does not affect the binocular disparities received by observers, we reasoned that subjects in our experiment would not believe that there is a potential dependency between transverse motion and disparity and would, therefore, be unable to learn that the disparity signal is also a cue to motion direction in our novel experimental environment.

Methods

Subjects

Subjects were eight students at the University of Rochester with normal or corrected-to-normal vision. All subjects were naive to the purposes of the study.

Stimuli

Motion stimuli were RDKs identical to those used in Experiment 1, except that, to limit "bleeding" across image frames in the stereo condition, only the red gun of the CRT was used. Stimuli containing binocular disparities were created as follows. Stationary dots were placed at simulated depths (all dots in a given display were at the same depth) ranging from −23 to 23 cm relative to fixation (or from 127 to 173 cm in absolute depth from the observer) and rendered from left-eye and right-eye viewpoints. Left-eye and right-eye images were presented to subjects using LCD shutter glasses (CrystalEyes 3 from Stereographics). Stimuli with both visual motion and disparity signals were created by placing moving dots at simulated depths and rendering the dots from left-eye and right-eye viewpoints.

Procedure

The procedure for Experiment 2 was identical to that for Experiment 1, except that the auditory signal was replaced by the binocular disparity signal. That is, subjects performed motion-only, disparity-only, and motion–disparity training trials, and motion–disparity test trials. For the motion–disparity training trials, stimuli with both motion and disparity signals were constructed as in Experiment 1, by mapping motion direction values onto disparity values based on the motion and disparity discrimination thresholds obtained in the motion-only and disparity-only training trials. The motion–disparity test trials were functionally identical to those of Experiment 1, with δ now representing offsets from the vertical direction in the disparity signal of the standard stimulus.

Results

If a subject had different PSEs when the disparity offset was $-\delta$ versus δ, then we can conclude that the subject learned to use the disparity signal as a cue to motion direction. Only one of the eight subjects had significantly different PSEs in the two conditions (at the $p < .05$ level), suggesting that subjects did not learn to use the disparity signal when judging motion direction.

The data for all subjects are shown in Figure 3. We fit psychometric functions (cumulative Normal distributions) to the combined data from all eight subjects when the offset in the standard was δ (solid line) and when it was $-\delta$ (dotted line). The average value across subjects for the offset δ was equivalent to a 4.49° rotation in motion direction. The experimental outcome is that subjects did not learn to use the disparity signal as a cue to motion direction—the 0.04° shift in PSEs when the offset was δ versus $-\delta$ was not significantly different from zero at the $p < .05$ level.

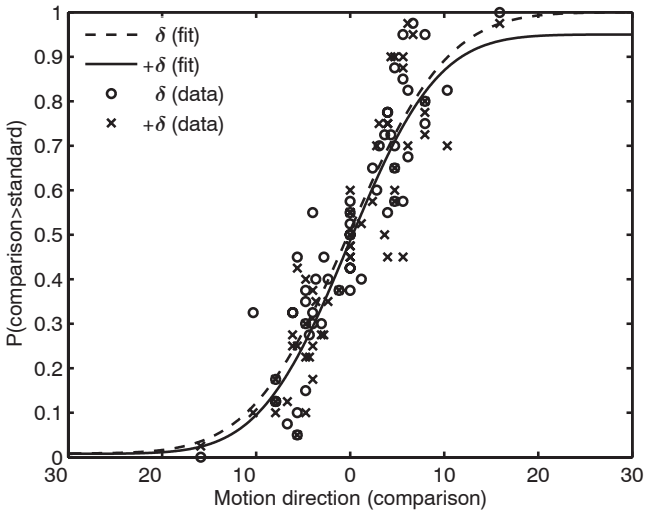

Figure 3. Data from the motion–disparity test trials for all eight subjects combined.

Experiment 3: Brightness cue to motion direction

Subjects in Experiment 3 were trained to perceive the direction of moving dots when the visual cue to motion direction was correlated with a visual brightness signal. The experiment examined whether subjects would learn that the brightness signal too is a cue to motion direction. Because the transverse motion of objects in the natural world does not affect their brightness, we reasoned that subjects do not represent a potential dependency between transverse motion and brightness and would, therefore, be unable to learn that the brightness signal is also a cue to motion direction in our novel experimental environment.

Methods

Subjects

Subjects were eight students at the University of Rochester with normal or corrected-to-normal vision. All subjects were naive to the purposes of the study.

Stimuli

Motion stimuli were RDKs identical to those used in Experiments 1 and 2, except that the individual dots were

assigned a neutral or pedestal brightness value. The brightness stimuli consisted of stationary random-dot images whose dots all shared a common pixel brightness value that ranged from 78 to 250 on a scale of 0–255. The pedestal pixel brightness of 164 had a luminance of 45.0 cd/m^2. Near this pedestal, luminance values scaled approximately linearly with pixel brightness, with 1 unit of RGB pixel brightness equivalent to 0.786 cd/m^2. Stimuli with both visual motion and brightness signals were created by assigning brightness pixel values to moving dots.

Procedure

The procedure for Experiment 3 was identical to those for Experiments 1 and 2, except that the auditory or disparity signals were replaced by a brightness signal.

Results

The motion–brightness test trials contained two conditions—the direction of the brightness signal in the standard stimulus was offset from vertical by an amount $-\delta$ or δ. If a subject had different PSEs in the two conditions, then we can conclude that the subject learned to use the brightness signal as a cue to motion direction. None of the eight subjects had significantly different PSEs in the two conditions (at the $p < .05$ level), suggesting that subjects did not learn to use the brightness signal when judging motion direction.

The data for all subjects are illustrated in Figure 4. We fit psychometric functions (cumulative Normal distributions) to the combined data from all eight subjects when the offset in the standard was δ (solid line) and when it was $-\delta$ (dotted line). The average value across subjects for the offset δ was equivalent to a 5.55° rotation in motion direction. The 0.70° shift in PSEs in the δ versus $-\delta$ cases was not statistically significant at the $p < .05$ level, indicating that subjects did not learn to use the brightness signal as a cue to motion direction.

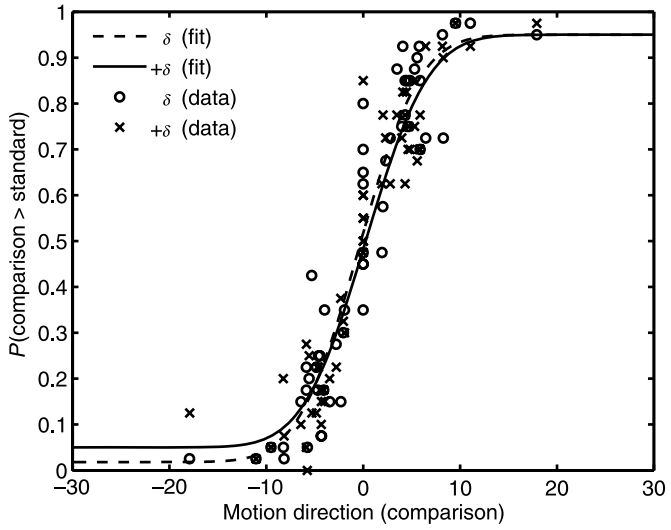

Figure 4. Data from the motion–brightness test trials for all eight subjects combined.

Discussion

We hypothesize that our early perceptual systems are capable of learning novel statistical relationships among scene and sensory variables that are already considered to be potentially dependent but that they cannot learn new relationships among scene and sensory variables that are not considered to be potentially dependent, even when placed in novel environments in which these variables are strongly related. Our experiments were designed to evaluate this hypothesis in the context of cue acquisition. Experiments 1, 2, and 3 evaluated whether people could learn new cues to transverse motion (motion in the frontoparallel plane).

In Experiment 1, subjects were exposed to an environment in which visual motion direction was correlated with an auditory signal. Because motion in natural environments often gives rise to both visual and auditory signals, it seems reasonable to assume that people believe that there is a potential dependency between motion direction and an auditory stimulus, and thus, we predicted that subjects would succeed in acquiring a new cue. The experimental results are consistent with this prediction. We can regard Experiment 1 as a control experiment—it establishes that our experimental procedures are adequate for inducing cue acquisition and that our statistical analyses are adequate for detecting this acquisition.

Experiments 2 and 3 exposed subjects to an environment in which visual motion direction was correlated with a binocular disparity signal or a brightness signal, respectively. In contrast to Experiment 1, cue acquisition in these cases requires representing statistical relationships among variables that do not share dependencies in the natural world. Transverse motion in natural environments does not lead to changes in disparity or brightness, and thus, people should not believe that there is a potential dependency between motion direction and disparity or brightness. We predicted that subjects would not acquire new cues to motion direction in these experiments, and the experimental results are consistent with these predictions.

There are at least two alternative explanations of our experimental results, however, that should be considered. First, perhaps there is something idiosyncratic about judgments of transverse motion. If so, one would not expect the experimental results to generalize to other perceptual judgments. Second, Experiment 1, where cue acquisition

was successful, used signals from different sensory modalities, whereas Experiments 2 and 3, where cue acquisition was not successful, used signals from a single modality. Perhaps this difference accounts for the differences in experimental outcomes. Experiments 4 and 5 were designed to evaluate these alternative explanations.

Experiment 4: Disparity cue to light source direction

Subjects in this experiment were trained to perceive the direction of a light source when the visual cue to light source direction—the pattern of shading across the visual objects—was correlated with a visual disparity signal. The experiment examined whether subjects would learn that the disparity signal is also a cue to light source direction. Because the direction of a light source has no effect on the depth of a lit object in the natural world, we reasoned that subjects should not represent a potential dependency between light source direction and disparity. Thus, we predicted that subjects would be unable to learn that the disparity signal is also a cue to light source direction in our novel experimental environment.

Methods

Subjects

Subjects were eight students at the University of Rochester with normal or corrected-to-normal vision. All subjects were naive to the purposes of the study.

Stimuli

Figure 5 depicts the stimuli used in Experiment 4. The shading stimuli consisted of 23 bumps (hemispheres) lying on a common frontoparallel plane whose pattern of shading provided information about the light source direction. Each bump subtended approximately 26 min of visual angle, and the bumps were scattered uniformly within a circular aperture (with a diameter of 6.28°). The light source was rendered as an infinite point source located 45° away from the frontoparallel plane along the z-axis (in the direction of the observer). The angular location of the light source varied from −90° (light coming from the left) to 90° (light coming from the right), with the light source direction in the standard stimulus always set to vertical (0°). In the shading-only training task, subjects viewed the stimuli monocularly with their dominant eyes. In all conditions, the bumps were rendered using only the red gun of the CRT.

The stimuli with binocular disparities were identical to those in the shading-only training task, except that the bumps were rendered from left-eye and right-eye viewpoints with

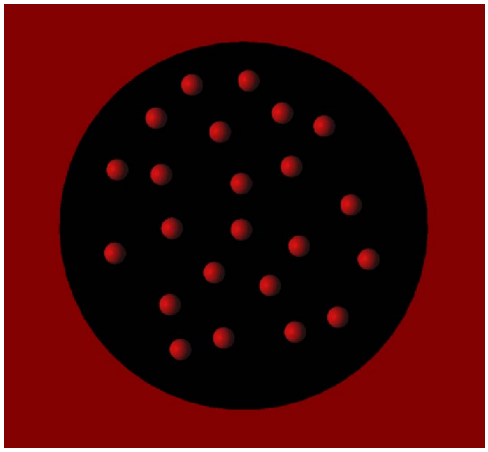

Figure 5. A sample stimulus from Experiment 4. In this example, the bumps are illuminated from the left.

flat lighting so that they appeared as discs of uniform luminance and, as with the static dots in Experiment 2, the discs were placed at simulated depths ranging from −23 to 23 cm relative to the observer (with all discs in a given display lying at a common depth). Stimuli with both shading and disparity signals were created by rendering the shaded bumps at simulated depths. In all tasks, each stimulus was presented for 1 s.

Procedure

The procedure for Experiment 4 was analogous to those for Experiments 1, 2, and 3. We used shading-only and disparity-only training tasks to characterize each subject's performance on lighting direction and depth discrimination tasks, respectively, and then trained subjects during shading–disparity training trials by exposing them to an environment in which disparity was correlated with shading. Finally, we tested subjects during shading–disparity test trials to evaluate whether they had learned that the disparity signal is also a cue to light source direction.

Results

The shading–disparity test trials contained two conditions—the direction of the disparity signal in the standard stimulus was offset from vertical by an amount $-\delta$ or δ. If a subject had different PSEs in the two conditions, then we can conclude that the subject learned to use the disparity signal as a cue to light source direction. None of the eight subjects had significantly different PSEs in the two conditions (at the $p < .05$ level), suggesting that subjects did not learn to use the disparity signal when judging light source direction.

The data for all subjects are illustrated in Figure 6. We fit psychometric functions (cumulative Normal distributions) to the combined data from all eight subjects when the

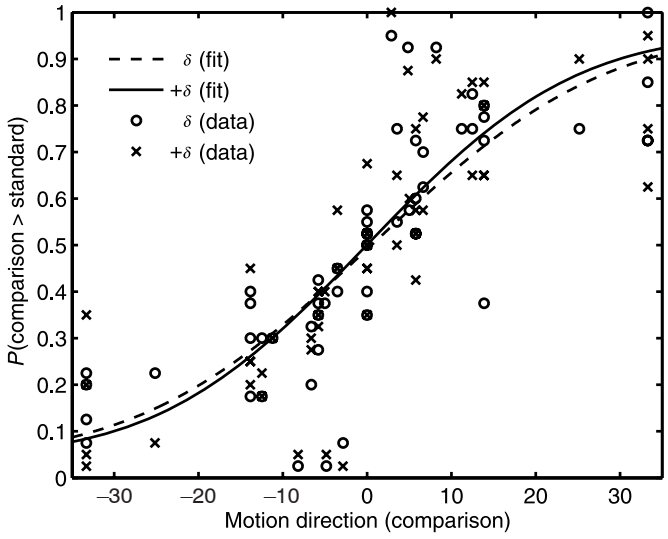

Figure 6. Data from the shading–disparity test trials for all eight subjects combined.

offset in the standard was δ (solid line) and when it was $-\delta$ (dotted line). The average value across subjects for the offset δ was equivalent to a 21.32° rotation in light source direction. The 0.59° shift in PSEs in the δ versus $-\delta$ cases was not statistically significant at the $p < .05$ level, indicating that subjects did not learn to use the disparity signal as a cue to light source direction.

Experiment 5: Auditory cue to light source direction

Subjects in Experiment 5 were trained to perceive the direction of a light source when the visual cue to light source direction—the pattern of shading across the visual objects—was correlated with a dynamic auditory signal. The experiment examined whether subjects would learn that the auditory signal is also a cue to light source direction. Because the direction of a light source has no effect on the motion of an object in the natural world—and thus, no effect on auditory signals—we reasoned that subjects should not represent a dependency between light source direction and the auditory signal. Thus, we predicted that subjects would be unable to learn that the disparity signal is also a cue to light source direction in our novel experimental environment.

Methods

Subjects

Subjects were eight students at the University of Rochester with normal or corrected-to-normal vision. All subjects were naive to the purposes of the study.

Stimuli

Shading stimuli consisted of 23 bumps (hemispheres) lying on a common frontoparallel plane whose pattern of shading provided information about the light source direction. Each bump subtended approximately 26 min of visual angle, and the bumps were scattered uniformly within a circular aperture (with a diameter of 6.28°). The light source was rendered as a diffuse panel source (i.e., as an array of local point sources) located 45° away from the frontoparallel plane along the z-axis (in the direction of the observer) and with its surface normal pointing toward the center of the bump array. The angular location of the light source varied from −90° (light coming from the left) to 90° (light coming from the right), with the light source direction in the standard stimulus always set to vertical (0°). Because we were concerned that subjects might be unable to bind the dynamic auditory signal with static visual stimuli in the combined trials, we changed the visual stimuli by jittering the light source so that the temporal microstructure of the visual stimulus seemed consistent with the dynamic (white noise) auditory signal. One of the point light sources in the panel array was selected at random and turned off in each frame to jitter the stimulus. This resulted in both flicker and positional jitter.

Auditory stimuli used in the auditory-only and shading–auditory trials were identical to those used in Experiment 1.

Procedure

The procedure for Experiment 5 was analogous to those for Experiments 1, 2, 3, and 4. We used shading-only and auditory-only training tasks to characterize each subject's performance on lighting direction and auditory discrimination tasks, respectively, and then trained the subjects during shading–auditory training trials by exposing them to an environment in which an auditory signal was correlated with shading. Finally, we tested subjects during shading–auditory test trials to evaluate whether they had learned that the auditory signal is also a cue to light source direction.

Results

The shading–auditory test trials contained two conditions—the direction of the auditory signal in the standard stimulus was offset from vertical by an amount $-\delta$ or δ. If a subject had different PSEs in the two conditions, then we can conclude that the subject learned to use the auditory signal as a cue to light source direction. None of the eight subjects had significantly different PSEs in the two conditions (at the $p < .05$ level), suggesting that subjects did not learn to use the auditory signal when judging light source direction.

The data for all subjects are illustrated in Figure 7. We fit psychometric functions (cumulative Normal distributions)

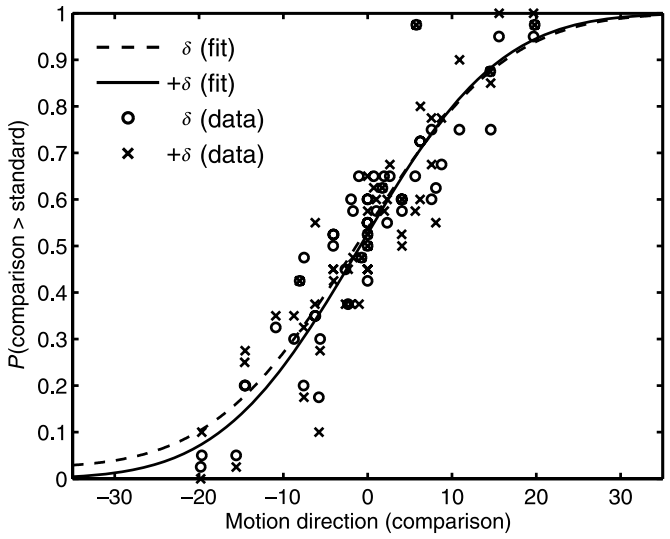

Figure 7. Data from the shading–auditory test trials for all eight subjects combined.

to the combined data from all eight subjects when the offset in the standard was δ (solid line) and when it was $-\delta$ (dotted line). The average value across subjects for the offset δ was equivalent to a 7.97° rotation in light source direction. The 0.073° shift in PSEs in the δ versus $-\delta$ cases was not statistically significant at the $p < .05$ level, indicating that subjects did not learn to use the auditory signal as a cue to light source direction.

General discussion

Numerous studies have shown that organisms' learning processes are often biased or constrained. This article has demonstrated that, similar to other learning processes, perceptual learning is biased, and has proposed a new constraint on early perceptual learning to account for this bias, namely, that people can modify their knowledge of the prior probabilities of scene variables or of the statistical relationships among scene and perceptual variables that are already considered to be potentially dependent but they cannot learn new relationships among variables that are not considered to be potentially dependent. An important goal of this article is to formalize these ideas using the notation of Bayesian networks and to illustrate how previous studies of early perceptual learning can be viewed as instances of parameter learning.

Bayesian networks are a tool for representing probabilistic knowledge that has been developed in the artificial-intelligence community (Neapolitan, 2004; Pearl, 1988). They have proven useful for modeling many aspects of machine and biological visual perception (e.g., Freeman, Pasztor, & Carmichael, 2000; Kersten, Mamassian, & Yuille, 2004; Kersten & Yuille, 2003; Schrater & Kersten, 2000). The basic idea underlying Bayesian networks is that a joint distribution over a set of random variables can be represented by a directed acyclic graph in which nodes correspond to variables and edges between nodes correspond to direct statistical dependencies among variables. For example, an edge from node x_1 to node x_2 means that the distribution of variable x_2 depends on the value of variable x_1 (as a matter of terminology, node x_1 is referred to as the parent of x_2). A Bayesian network is a representation of the following factorization of a joint distribution:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | pa(x_i)), \quad (2)$$

where $P(x_1, \ldots, x_n)$ is the joint distribution of variables $x_1, \ldots, x_n$ and $P(x_i \mid pa(x_i))$ is the conditional distribution of x_i given the values of its parents [if x_i has no parents, then $P(x_i \mid pa(x_i)) = P(x_i)$].

As an introduction to Bayesian networks, consider the following scenario: an observer pulls into a parking lot, and as she begins to exit her car, she hears a Rotweiller's bark. Looking in the direction of the bark, she sees a distant Rotweiller. The observer's car is parked close to a building's entrance, and the observer must decide whether to wait for the dog to leave the vicinity or to try to make it to the entrance before encountering the dog. In making this decision, the observer would like to know the following: (1) Is the dog dangerous? and (2) How far away is the dog? To simplify things, assume that the observer has access to only three pieces of information: the loudness of the dog's bark and the size of the dog's image, which are both cues to the distance of the dog, and whether the dog is foaming at the mouth, which lets the observer know whether the dog is rabid and, therefore, dangerous (for simplicity, assume that only rabid dogs are dangerous).

Figure 8 shows a Bayesian network that represents this situation. The variables corresponding to scene properties are located toward the top of this figure, whereas the variables corresponding to percepts are located toward the bottom. Scene variables do not have parents, although they serve as parents to sensory variables as indicated by the arrows. A Bayesian network is a useful representation of the joint distribution of scene and sensory variables because of the way it represents potential dependencies. Although statistical dependency and causality are not equivalent relationships, Bayesian networks are often interpreted as instances of "generative models" whose edges point in the direction of causality. Consider the edges in Figure 8. A change in the physical distance from the dog to the observer causes a change in the perceived size of the dog's image on the observer's retina and in the perceived loudness of the dog's bark at the observer's ear. Likewise, rabies may lead to the observer perceiving the dog to foam at the mouth. These relationships, however, are not deterministic; the perceived size of the dog's

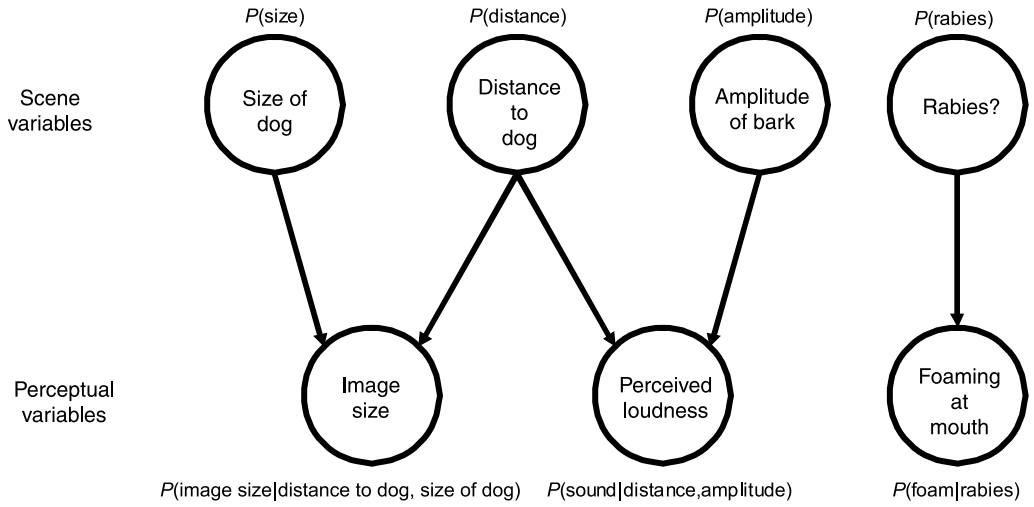

Figure 8. A simple Bayesian network representing a distant barking dog that may or may not have rabies. The variables corresponding to scene properties are located toward the top of this figure, whereas the variables corresponding to percepts are located toward the bottom. Scene variables do not have parents, although they serve as parents to sensory variables as indicated by the arrows.

image and the perceived loudness of the dog's bark can also vary due to additional factors that are difficult to measure, such as physical or neural noise. The conditional distributions associated with sensory variables represent these uncertainties.

Bayesian networks are most useful when they represent relationships among variables in ways that are both sparse and decomposable.[4] The structure of the graph in Figure 8 has been greatly simplified using our knowledge about causality, and thus, the graph represents potential relationships among variables in a sparse way. We understand, for example, that knowing that the dog has rabies or is foaming at the mouth tells us nothing about the dog's distance, its retinal image size, or the loudness of its bark. Consequently, there are no edges that link the former and latter variables. It is precisely these sorts of simplifications, or assumed independencies, that make reasoning computationally feasible. For example, an observer reasoning about whether a dog has rabies only needs to consider whether the dog is foaming at the mouth and can ignore the values of all other variables. Bayesian networks also represent relationships in ways that are decomposable. An observer wishing to estimate the distance to a dog based on the dog's retinal image size and the loudness of the dog's bark can do so using Bayes' rule:

$$P(\text{distance to dog} \mid \text{image size, loudness of bark}) \propto P(\text{image size, loudness of bark} \mid \text{distance to dog}) \times P(\text{distance to dog}). \quad (3)$$

This calculation can be simplified by noting that, according to the Bayesian network in Figure 8, the size of the dog's image and the loudness of its bark are conditionally independent given the distance to the dog. Consequently, the joint distribution on the right-hand side of Equation 3 can be factored as follows:

$$P(\text{image size, loudness of bark} \mid \text{distance to dog}) = P(\text{image size} \mid \text{distance to dog}) \times P(\text{loudness of bark} \mid \text{distance to dog}). \quad (4)$$

The computational advantages of statistical relationships that are sparse and decomposable are difficult to appreciate in a simple scenario with seven random variables but have enormous importance in real-world situations with hundreds of variables. Indeed, whether reasoning requires the consideration of only a small subset of variables versus the need to take into account all variables or whether reasoning requires the calculation of high-dimensional joint distributions versus low-dimensional distributions are typically the most important factors that make a problem solvable versus unsolvable in practice (Bishop, 1995; Neapolitan, 2004).

Using the notation of Bayesian networks, we can restate our hypothesis about constraints on early perceptual learning. Recall our hypothesis that people's early perceptual processes can modify their knowledge of the prior probabilities of scene properties or of the statistical relationships among scene and sensory variables that are already considered to be potentially dependent. However, they cannot learn new relationships among scene and sensory variables that are not considered to be potentially dependent. In terms of Bayesian networks, our hypothesis states that early perceptual processes can modify their prior probability distributions for scene variables or their conditional probability distributions, specifying how sensory variables depend on scene variables. However, they cannot add new nodes or new edges between scene and sensory variables in their graphical representation. In the machine

learning literature, researchers make a distinction between "parameter learning," which means learning the prior and conditional probability distributions, and "structure learning," which means learning the nodes and edges of the graphical structure. Using the terminology of this literature, our hypothesis states that early perceptual processes are capable of parameter learning, but they are not capable of structure learning.[5] Interestingly, parameter learning is often thought to be computationally feasible—the machine learning literature contains several maximum-likelihood and Bayesian algorithms that often work well in practice. In contrast, structure learning is widely regarded as intractable—there are currently no general-purpose algorithms for structure learning that work well on moderate- or large-sized problems (Rish, 2000).[6]

Our hypothesis can be divided into two parts: (i) early perceptual processes are capable of parameter learning, and (ii) early perceptual processes are not capable of structure learning. To our knowledge, there are no demonstrations in the scientific literature of structure learning by early perceptual processes. That is, we believe that all examples of early perceptual learning are demonstrations of parameter learning. To illustrate this point, we review several classes of early perceptual learning phenomena.

Many studies on perceptual learning report the results of experiments in which observers show improved performance on tasks requiring fine discriminations along simple perceptual dimensions. For example, observers might improve at tasks that require the discrimination of motion directions of single dots (Matthews & Welch, 1997) or fields of dots (Ball & Sekuler, 1987), of orientations of line segments (Matthews & Welch, 1997), of spatial frequencies within plaid patterns (Fine & Jacobs, 2000), or of offsets between nearly collinear line segments (Fahle, Edelman, & Poggio, 1995). Often, the learning demonstrated using such tasks is stimulus specific in the sense that the learning fails to generalize to novel versions of the task using different stimulus positions or configurations.

Figure 9 shows a Bayesian network depicting the dependence assumptions that might underlie performance on a task that requires observers to make fine discriminations along a simple perceptual dimension. In this figure, the physical scene gives rise to a set of conditionally independent perceptual features that the observer uses to make decisions regarding the value of some scene property. An account of how an observer improves at estimating the scene property based on the perceptual features is as follows. The process of learning consists of improving the estimates of the relationships between the scene property and the values of each of the perceptual features. In terms of Bayesian networks, learning consists of improving the estimates of the conditional distributions associated with the perceptual variables, that is, the distributions of the values of the features given a value (or an estimated value) of the scene property: $P(\text{feature } i \mid \text{scene} = x)$ for $i = 1, \ldots, N$. Of particular interest are the variances of these distributions because these variances suggest how reliable or informative each feature is in indicating the value of the scene property. Features whose values have a large variance for a fixed value of the scene property are relatively unreliable or uninformative. In contrast, features whose values have a small variance for a fixed scene value are more reliable. More accurate estimates of the conditional distributions and, thus, their variances allow an observer to more accurately weight features according to their relative reliabilities, effectively placing greater weight on more reliable features and lesser weight on less reliable features. Most important, for our purposes, observers' improved performances can be accounted for solely based on parameter learning; structure learning is not required.

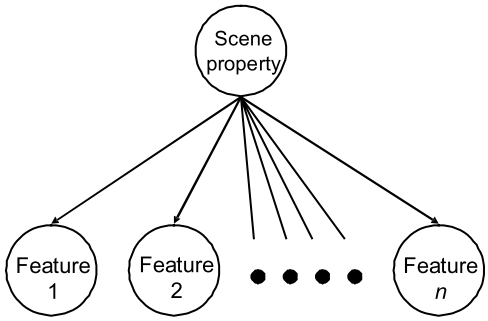

Figure 9. A Bayesian network depicting the dependence assumptions underlying perceptual judgments in tasks that require observers to make a fine discrimination along a simple perceptual dimension.

Ball and Sekuler (1987) reported an experiment in which observers improved at discriminating motion directions in RDKs centered at a training direction but did not show improved performance when tested with motion directions centered at a novel orthogonal direction. An account of this finding based on Bayesian networks is as follows. Consider a network in which the parent node corresponds to a scene variable representing the direction of motion in the scene and in which the N child nodes correspond to sensory feature detectors sensitive to image motion in N different directions. When observers view a display of a kinematogram containing a motion direction near the training direction, they first estimate this direction based on the values of their feature detectors. They then adapt their estimates of the conditional distributions of the values of the feature detectors given the estimated value of the motion direction. Over the course of many trials, observers learn which feature detectors have a small variance for a given value of the motion direction, which means that they learn which feature detectors are most reliable for judging motion direction for directions near the training direction.[7] Because observers do not modify their distributions that are conditioned on other motion directions, they do not show improved performance when tested with kinematograms whose motion directions are centered at a novel orthogonal direction.

A demonstration that observers learn to weight features according to their relative reliabilities in a manner consistent with the account of learning described above was provided by Gold, Sekuler, and Bennett (2004). These researchers examined observers' perceptual representations through the construction of "classification images." Briefly, classification images are created by correlating image feature values (e.g., pixel luminances) with the values of a scene property. Image features that vary reliably with the scene property take on extreme values in the classification image, whereas unreliable features take on values near zero. Classification images are often used in the context of perceptual classification. An ideal classification image for a set of stimuli is constructed by correlating the feature values of each stimulus with its correct classification; classification images for individual observers can be created by correlating the feature values of each stimulus with the classification indicated by the observer. For difficult tasks, the ideal classification images tend to be relatively sparse, with few reliable features for discriminating between stimulus classes. Our account of learning described above predicts that naive observers should initially use a large set of features when discriminating stimuli and then gradually reduce the influence of many features as they discover which features are most reliable for the task. Gold et al.'s results suggest that, during the course of learning, observers' classification images move toward the ideal classification image in exactly this manner, with observers incrementally basing their decisions on a smaller, more reliable subset of the available features.

A second class of phenomena studied in the perceptual learning literature is the acquisition of new cue combination rules. Several studies have found that observers modify how they combine information from two visual cues when one of the cues is made less reliable (Atkins, Fiser, & Jacobs, 2001; Ernst, Banks, & Bülthoff, 2000; Jacobs & Fine, 1999). Ernst et al. (2000), for example, placed observers in an environment in which the slant of a surface indicated by a haptic cue was correlated with the slant indicated by one visual cue but uncorrelated with the slant indicated by another visual cue (this slant value varied randomly over trials). Observers adapted their visual cue combination rules to place more weight on the information derived from the visual cue that was consistent with haptics and less weight on the information derived from the other visual cue. This class of learning phenomena can be regarded as conceptually equivalent to the first class of phenomena described above in which observers modify how they combine information from multiple feature detectors. Consequently, our account of learning for this second class is very similar to our account for the first class.

The top node of the Bayesian network in Figure 10 represents a scene variable, such as the slant of a surface, whereas the two child nodes represent corresponding sensory variables based on two perceptual cues, such as slant from visual stereo and slant from visual texture. Imagine that both cues are normally good indicators of the

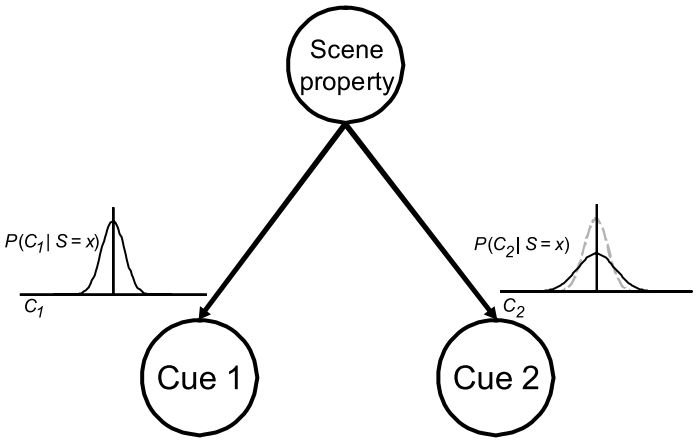

Figure 10. A Bayesian network representing the type of modification that might underlie the acquisition of new cue combination rules. Cue 1 represents a cue whose reliability is fixed, whereas Cue 2 represents a cue that has become less reliable. The solid black curves represent the final conditional cue distributions given a value (or estimated value) of the scene property. The dashed gray curve represents the conditional distribution for Cue 2 before learning.

scene property, but the observer is placed in a novel environment in which Cue 1 is reliable but Cue 2 is not [e.g., the variance of $P(\text{Cue 1} \mid \text{scene} = x)$ is small, whereas the variance of $P(\text{Cue 2} \mid \text{scene} = x)$ is large]. An account of how an observer improves at estimating the scene property based on the two perceptual cues is as follows. On each trial of an experiment, observers first estimate the value of the scene property based on the values of all sensory variables. They then improve the estimates of the relationships between the scene property and the values of each of the cues; that is, observers modify their estimates of $P(\text{Cue 1} \mid \text{scene} = x)$ and $P(\text{Cue 2} \mid \text{scene} = x)$, where x is the estimated value of the scene property. More accurate estimates of these distributions, particularly their variances, allow an observer to more accurately weight cues according to their relative reliabilities, effectively placing greater weight on more reliable cues and lesser weight on less reliable cues. As above, observers' improved performances can be accounted for solely based on parameter learning and does not require structure learning.

A third class of perceptual learning phenomena is often referred to as "cue recalibration" (e.g., Atkins, Jacobs, & Knill, 2003; Bedford, 1993; Epstein, 1975; Harris, 1965; Mather & Lackner, 1981; Welch, 1986). For example, an observer may wear prisms that shift the visual world 10° to the right. As a result, objects visually appear at locations 10° to the right of the locations indicated by other sensory signals. Over time, observers notice this discrepancy and recalibrate their interpretations of the visual cue so that the visual location is more consistent with the locations indicated by other sensory cues.

Using our Bayesian network framework, we hypothesize that observers first estimate the value of the scene property

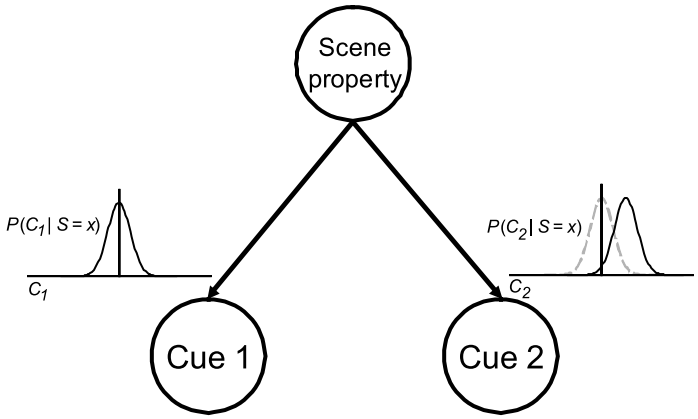

Figure 11. A Bayesian network representing the type of modification that might underlie perceptual recalibration. Cue 1 represents an accurate and low-variance cue, whereas Cue 2 represents a low-variance cue but whose estimates are no longer accurate. The solid curves represent the final conditional cue distributions for a particular value of the scene variable. The dashed gray curve represents the conditional distribution for Cue 2 before learning.

(top node of Figure 11) based on the values of all sensory cues (bottom nodes of Figure 11). They then modify their estimates of the conditional distributions associated with the sensory variables: $P(\text{Cue 1} \mid \text{scene} = x)$ and $P(\text{Cue 2} \mid \text{scene} = x)$, where x is the estimated value of the scene property. Unlike the case of learning new cue combination rules, the modification is not primarily to the estimate of the variance of the distribution associated with a newly unreliable cue. Rather, it is to the estimate of the mean of the distribution associated with a newly uncalibrated cue (due, perhaps, to the shift in the visual world caused by prisms). As before, observers' improved performances can be accounted for solely based on parameter learning.

The last class of perceptual learning phenomena that we consider here is the acquisition of visual priors. For example, consider observers viewing displays of circular patches that are lighter toward their top and darker toward their bottom. These displays are consistent with a bump that is lit from above or a dimple that is lit from below. Observers tend to assume that the light source is above a scene and, thus, prefer to interpret the object as a bump. Observers in an experiment by Adams, Graf, and Ernst (2004) viewed objects whose shapes were visually ambiguous and also touched these objects, thereby obtaining haptic information disambiguating the objects' shapes. The shape information obtained from haptics was consistent with an interpretation of the visual display in which the estimated light source location was offset from its expected location based on observers' prior probability distributions of the light source's location. Adams et al. found that observers modified their prior distributions to reduce the discrepancy between estimated and expected light source locations.

The Bayesian network in Figure 12 has two scene variables, corresponding to the object's shape and the light source's location, and a sensory variable, corresponding to the perceived visual shape of the object. Our account of learning in this setting is as follows. Based on an unambiguous haptic percept (not shown in Figure 12) and the ambiguous visual percept, observers estimate the object's shape. Based on this shape and the perceived visual shape, observers then estimate the light source's location. Learning occurs due to the discrepancy between the estimated location of the light source and the expected location based on observers' prior probability distribution of this location. To reduce this discrepancy, observers modify their prior distribution appropriately. Thus, as in the other classes of learning phenomena reviewed above, the acquisition of prior probability distributions can be accounted for through parameter learning and does not require structure learning.

We have reviewed four classes of early perceptual learning phenomena and outlined how they can be accounted for solely through parameter learning. We hypothesize that all early perceptual learning is parameter learning; that is, all early perceptual learning involves the modification of knowledge of the prior probabilities of scene properties or of the statistical relationships among scene and sensory variables that are already considered to be potentially dependent. Conversely, we hypothesize that early perceptual learning processes are biased or constrained such that they are incapable of structure learning

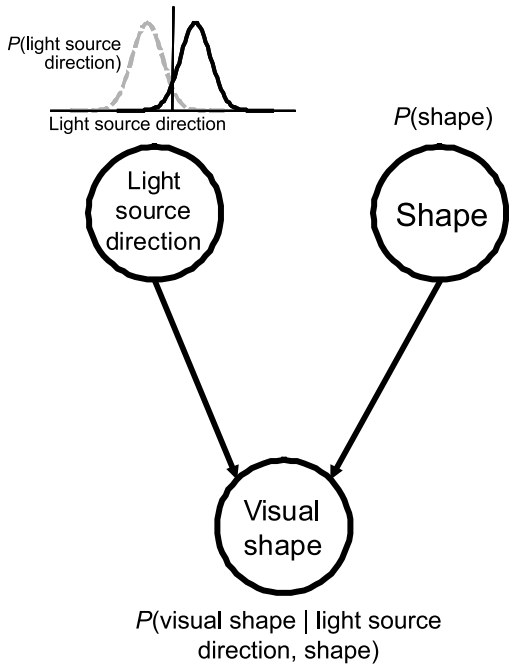

Figure 12. A Bayesian network characterizing subjects' modifications of their prior distribution of the light source location in the experiment reported by Adams et al. (2004).

(the addition of new nodes or new edges between scene and sensory variables), which means that these processes cannot learn new relationships among scene and sensory variables that are not considered to be potentially dependent.

In the experimental part of this article, we reported the results of five experiments that evaluate whether subjects can demonstrate cue acquisition. Figures 13 and 14 illustrate the relationships between scene and sensory variables in Experiments 1, 2, and 3 and Experiments 4 and 5, respectively, in terms of Bayesian networks. Here, the solid black edges represent dependencies that exist in the natural world, whereas the dashed gray edges represent dependencies that do not exist in the natural world but that we introduced in our novel experimental environments. For the reasons outlined in the experimental sections, we expected that observers started our experiments with the belief that variables that are connected by a black edge are potentially dependent, whereas variables that are connected by a gray dashed edge are not.

In Experiment 1, subjects were placed in a novel environment that resembled natural environments in the sense that it contained systematic relationships among scene and perceptual variables that are normally dependent. In this case, cue acquisition requires parameter learning and, as predicted, subjects succeeded in learning a new cue. In Experiments 2, 3, 4, and 5, subjects were placed in novel environments that did not resemble natural environments—they contained systematic relationships among scene and perceptual variables that are not normally dependent. Cue acquisition requires structure learning in these cases. Consistent with our hypothesis, subjects failed to learn new cues in Experiments 2, 3, 4, and 5. Taken as a whole, our hypothesis provides a good account of the pattern of experimental results reported here. That is, it explains why people learn in some situations and fail to learn in other situations.

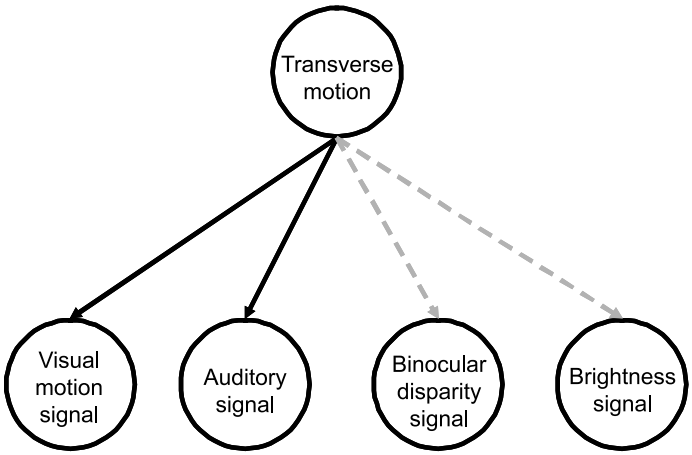

Figure 13. A Bayesian network representing the statistical relationships studied in Experiments 1, 2, and 3. The solid black edges represent dependencies that exist in the natural world, whereas dashed gray edges represent dependencies that do not exist in the natural world but that we introduced in our novel experimental environments. We expect that observers started our experiments with the belief that variables connected by a black edge are potentially dependent, whereas variables connected by a gray edge are not.

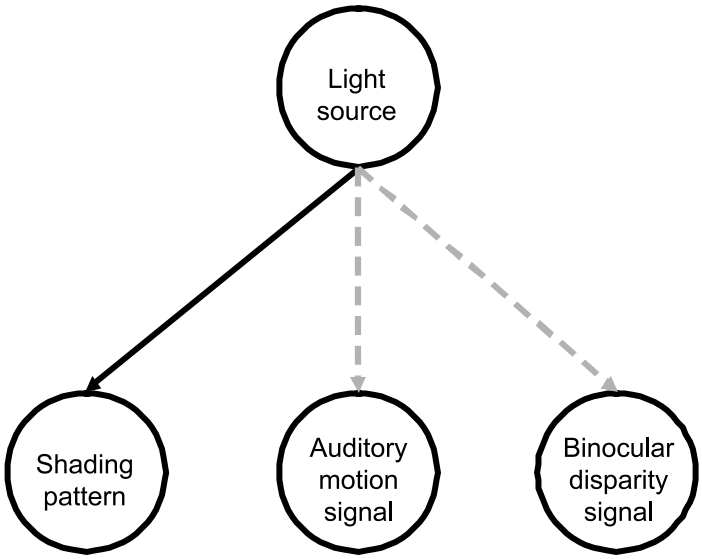

Figure 14. A Bayesian network representing the statistical relationships studied in Experiments 4 and 5. The solid black lines represent preexisting edges—conditional dependencies that exist in the natural world—whereas the dashed gray lines represent conditional dependencies that do not exist in the natural world but that we introduced in our novel experimental environments.

In addition to providing an account of experimental data, our hypothesis also has the property of being motivated by computational considerations. As discussed above, machine learning researchers have found that parameter learning in networks with sparse connectivity is a comparatively easy problem, whereas structure learning is typically intractable. Thus, there are good computational reasons why early perceptual learning processes might be constrained in the ways hypothesized here.

Our theory is limited to early perceptual learning and is not intended to be applied to late perceptual or cognitive learning. This point can be demonstrated in at least two ways. First, it seems reasonable to believe that learning to visually recognize an object involves structure learning. Gauthier and Tarr (1997), for example, trained subjects to visually recognize objects referred to as "greebles." A plausible account of what happens when a person learns to visually recognize a novel object as the greeble named "pimo" is that the person adds a new node (along with new edges) to his or her Bayesian network representation that corresponds to this newly familiar object. If this speculation is correct, then it raises the question of why structure learning is computationally feasible for late perceptual learning but intractable for early perceptual learning. It may be that structure learning of higher level knowledge becomes feasible when a preexisting structure representing lower level knowledge is already in place.

Second, there have been several demonstrations of "contextually dependent" perceptual learning that, we conjecture, may be accounted for via late perceptual learning processes performing structure learning. Atkins et al. (2001), for example, trained subjects to combine depth information from visual motion and texture cues in one way when the texture elements of an object were red and to combine information from these cues in a different way when the elements were blue. In other words, the discrete color of the elements signaled which of two contexts subjects were currently in, and these two contexts required subjects to use different cue combination rules to improve their performance on an experimental task. Because there is no systematic relationship between color and cue combination rule in natural environments, people should not believe that color and cue combination rule are potentially dependent variables, which means that the type of learning demonstrated here would require structure learning. Related results in the domain of cue acquisition have recently been reported by Haijiang et al. (2006). We speculate that this type of contextually dependent perceptual learning is due to learning processes that have a higher level than the processes that we have focused on in this article.[8]

We have described a hypothesis about constraints on early perceptual learning. Admittedly, the hypothesis is speculative. Although the data favoring the hypothesis are currently sparse, its advantages include the following: it accounts for an important subset of data about perceptual learning that would otherwise be confusing; it uses a Bayesian network formulation that is well specified (and, thus, falsifiable) and mathematically rigorous; and it leads to several theoretically interesting and important research questions. The hypothesis is meant to deal with perceptual learning in general, although the experiments in this article have focused on the predictions of our hypothesis for perceptual cue acquisition. For us, this seems to be a natural place to begin because the hypothesis's predictions with respect to cue acquisition are straightforward. Future work will focus on delineating and testing the hypothesis's predictions on other perceptual learning tasks. A primary challenge of such work will lie in developing efficient experimental methods for detecting changes (or the lack thereof) in observers' representations of variable dependencies.

Acknowledgments

We thank the reviewers for helpful comments on an earlier version of this manuscript. This work was supported by NIH research grant R01-EY13149.

Commercial relationships: none.
Corresponding author: Robert A. Jacobs.
Email: robbie@bcs.rochester.edu.
Address: 416 Meliora Hall, Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627-0268.

Footnotes

[1]We use the terms "low-level perception," "early perception," or "early perceptual learning" in the same ways as many other researchers in the perceptual sciences literature (see Fahle & Poggio, 2002; Gilbert, 1994, for reviews). Although the exact meanings of these terms can be fuzzy—for example, the boundary between early versus late perception is not completely understood—investigators have found these terms to be highly useful.

[2]The visual and auditory stimuli in this experiment are analogous to small particles (e.g., sand grains) moving across a surface with an anisotropic texture. The sound produced by such a stimulus depends on the properties of the surface texture, and, as in the current experiment, the anisotropic surface texture would cause changes in the mean direction of the moving particles to lead to systematic changes in spectral properties of the resulting sound.

[3]A reader may wonder why we did not use more traditional auditory motion stimuli with interaural phase and intensity differences. There are two reasons for this. First, we wanted to make the auditory signal ambiguous; setting up systematic interaural differences would bias observers to perceive the auditory stimulus as indicating a particular motion direction. Second, the visual stimuli used in the current experiment represent *field* motion and not *object* motion (i.e., the mean velocity of dot fields in our RDKs varies between stimuli, whereas the mean position of these dots remains constant). Interaural phase and intensity differences result from changes in the position of an object—or in the mean position of a group of objects. Because the mean positions of our visual stimuli remain constant, interaural differences are inappropriate for representing their motion.

[4]A reader might wonder why a fully connected Bayesian network (i.e., one in which all scene variables connect to all sensory variables) is not always used. An advantage of such a network is that it could represent any relationship between scene and sensory variables. Unfortunately, as mentioned in this article and detailed in the literature on machine learning, there is a price to pay for such representational richness—inference and learning in fully connected networks are prohibitively expensive in terms of computation. In fact, inference and learning are computationally feasible only in networks with sparse connectivity. Similarly, a reader might wonder why a network that initially contains no connections but in which connections are added over time as needed is not always used. As before, this type of "structure learning" is prohibitively expensive from a computational viewpoint.

[5]If people's early perceptual knowledge can be characterized by Bayesian networks but the structures of these networks are not learned, then this raises the question of where these structures come from. We speculate that people's network structures are innate, resulting from our

evolutionary history in an environment with stationary physical laws. If this "strong" view is not strictly correct, we would not be uncomfortable with a "weaker" view in which the structures are fixed in adults but that structural learning takes place during infancy or early childhood.

[6]As a technical detail, it is worth noting that both parameter and structure learning in Bayesian networks are often regarded as NP-hard problems (Cooper, 1990; Jordan & Weiss, 2002). To illustrate why parameter learning is regarded as NP-hard, one must keep in mind that inference (determining the conditional distribution of some variables given the values of other variables) is often a subproblem of parameter learning (e.g., the E-step of the Expectation–Maximization [EM] algorithm requires inference). In the machine learning community, inference is typically performed using the junction-tree algorithm. The computational complexity of this algorithm is a function of the size of the cliques upon which message-passing operations are performed. Unfortunately, summing a clique potential is exponential in the number of nodes in the clique. This fact motivates the need to use Bayesian networks with sparse connectivity. Because such networks tend to minimize clique sizes, inference (and, thus, parameter learning) in these networks is often feasible. To illustrate why structure learning is regarded as NP-hard, one must keep in mind that structure learning is typically posed as a model selection problem within a hierarchical Bayesian framework. The top level of the hierarchy includes binary variables $\{M_i, i = 1, \ldots, K\}$, where M_i indicates whether model i is the "correct" model and the number of such models K is superexponential in the number of nodes in the network. The middle level includes real-valued variables $\{\theta_i, i = 1, \ldots, N\}$, where θ_i is the set of parameters for model i. The bottom level is the data, denoted D. The likelihood for model i, $P(D \mid M_i)$, is computed as follows: $P(D \mid M_i) = \int P(D \mid \theta_i)\, P(\theta_i \mid M_i)\, d\theta_i$. Note that $P(D \mid \theta_i)$ is the likelihood for the parameters θ_i (used during parameter learning). Also note that this integral is typically not analytically solvable.

[7]Readers familiar with the machine learning literature will recognize that we are conjecturing that the observer's learning rule resembles an EM algorithm, an algorithm for maximizing likelihood functions. At first blush, it may seem that the observer is faced with a "chicken-and-egg" problem: Observers first use their feature detectors to estimate the motion direction and then use the estimated motion direction to determine the most reliable features. The EM algorithm is often used to solve such chicken-and-egg problems (see Dempster, Laird, & Rubin, 1977).

[8]Interestingly, contextually dependent learning is often regarded as different than most other forms of learning. For example, researchers in the animal learning theory community distinguish standard forms of learning, which they refer to as associative learning, from contextually dependent learning, which they refer to as occasion setting (Schmajuk & Holland, 1998).

References

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7,* 1057–1058. [PubMed] [Article]

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research, 41,* 449–461. [PubMed]

Atkins, J. E., Jacobs, R. A., & Knill, D. C. (2003). Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts. *Vision Research, 43,* 2603–2613. [PubMed]

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

Bedford, F. (1993). Perceptual learning. *Psychology of Learning and Motivation, 30*, 1–60.

Bishop, C. M. (1995). Chapter 2: Probability density estimation. In C. M. Bishop (Ed.), *Neural networks for pattern recognition* (pp. 33–76). New York, NY: Oxford University Press.

Brunswick, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.

Cooper, G. F. (1990). The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence, 42*, 393–405.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B, 34,* 1–38.

Epstein, W. (1975). Recalibration by pairing: A process of perceptual learning. *Perception, 4*, 59–72. [PubMed]

Ernst, M. O., Banks, M. S., & Bülthoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience, 3,* 69–73. [PubMed] [Article]

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research, 35,* 3003–3013. [PubMed]

Fahle, M., & Poggio, T. (2002). *Perceptual learning*. Cambridge, MA: MIT Press.

Fine, I., & Jacobs, R. A. (2000). Perceptual learning for a pattern discrimination task. *Vision Research, 40,* 3209–3230. [PubMed]

Freeman, W. T., Pasztor, E. C., & Carmichael, O. T. (2000). Learning low-level vision. *International Journal of Computer Vision, 40,* 25–47.

Garcia, J., & Koelling, R. A. (1966). The relation of cue to consequence in avoidance learning. *Psychonomic Science, 4,* 123–124.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37,* 1673–1682. [PubMed]

Gilbert, C. D. (1994). Early perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America, 91,* 1195–1197. [PubMed] [Article]

Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science, 28,* 167 – 207.

Haijiang, Q., Saunders, J. A., Stone, R. W., & Backus, B. T. (2006). Demonstration of cue recruitment: Change in visual appearance by means of Pavlovian conditioning. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 483 – 488. [PubMed] [Article]

Harris, C. S. (1965). Perceptual adaptation to inverted, reversed, and displaced vision. *Psychological Review, 72,* 419–444. [PubMed]

Jacobs, R. A., & Fine, I. (1999). Experience-dependent integration of texture and motion cues to depth. *Vision Research, 39,* 4062– 4075. [PubMed]

Jordan, M. I., & Weiss, Y. (2002).Graphical models: Probabilistic inference. In M. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed.). Cambridge, MA: MIT Press.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55,* 271 –304. [PubMed]

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology, 13,* 150 –158. [PubMed]

Mather, J., & Lackner, J. R. (1981). Adaptation to visual displacement: Contribution of proprioceptive, visual, and attentional factors. *Perception, 10,* 367–374. [PubMed]

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

Neapolitan, R. E. (2004). *Learning Bayesian networks*. Upper Saddle River, NJ: Pearson Education.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufmann.

Rish, I. (2000). Advances in Bayesian learning. *Proceedings of the 2000 International Conference on Artificial Intelligence*. CSREA Press.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47,* 172 –196.

Schmajuk, N. A., & Holland, P. C. (1998). *Occasion setting*. Washington, DC: American Psychological Association.

Schrater, P. R., & Kersten, D. (2000). How optimal depth cue integration depends on the task. *International Journal of Computer Vision, 40,* 71– 89.

Wallach, H. (1985). Learned stimulation in space and motion perception. *American Psychologist, 40,* 399– 404. [PubMed]

Welch, R. B. (1986). Adaptation of space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (vol. 1, pp. 1–45). New York: Wiley-Interscience.

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63,* 1293–1313. [PubMed] [Article]

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63,* 1314–1329. [PubMed] [Article]