CHAPTER 4

Generative Probabilistic Modeling: Understanding Causal Sensorimotor Integration

Sethu Vijayakumar, Timothy Hospedales, and Adrian Haith

INTRODUCTION

In this chapter, we argue that many aspects of human perception are best explained by adopting a modeling approach in which experimental subjects are assumed to possess a full generative probabilistic model of the task they are faced with, and that they use this model to make inferences about their environment and act optimally given the information available to them. We apply this generative modeling framework in two diverse settings—concurrent sensory and motor adaptation, and multisensory oddity detection—and show, in both cases, that the data are best described by a full generative modeling approach.

Bayesian ideal-observer modeling is an elegant and successful normative approach to understanding human perception. One particular domain in which it has seen much success recently is that of understanding multisensory integration in human perception (see Chapter 1).

Existing applications of this modeling approach have frequently focused on a simple special case where the ideal observer's estimate of an unknown quantity in the environment is a reliability-weighted mean of the individual observed cues. This is all that is needed to understand a wide variety of interesting perceptual phenomena. We argue, however, that the Bayesian-observer approach can be more powerfully and generally applied by clear generative modeling of the perceptual task for each experiment. In other words, this assumes that people have access to a full generative model of their observations and that they use this model to make optimal decisions in performing the task.

This systematic approach effectively provides a "model for modeling" that has some key advantages: (1) It provides the modeler with a clear framework for modeling new tasks beyond simply applying common normative models—such as linear combination—which may not apply for a new scenario and may fail to explain important aspects of human behavior; (2) Human performance can be measured against these clear "optimal" models such that we can draw conclusions about optimality of human perception or reveal architectural limitations of the human perceptual system, which cause it to deviate from optimality.

For a particular perceptual task, the optimal solution requires inference in the true generative model of the task. Here, optimal is defined in the sense that the posterior probability over relevant unknowns in the environment is calculated. Any actions or decisions to be made can then be taken

with respect to this posterior and the required loss function (see Chapter 1). As an intuition for the significance of optimality, consider that someone gambling on the state of the real world given this "optimal" posterior is guaranteed not to lose money in the long term to someone gambling with any other distribution, including the posterior from a "wrong" generative model.

To make predictions about human behavior, the modeler must therefore take care to construct a generative model that encompasses all relevant aspects of the task. These models often lead to strong and surprising new predictions, which can be tested experimentally. In this chapter, we illustrate these ideas via two experiments for which we show that it is crucial to consider a complete normative generative model of the data. In these cases, naive application of common simple models fails to even qualitatively explain the data. Rather than conclude that human perception is suboptimal in these ways, we show how a full generative modeling approach can explain the data and provide insight into human behavior.

We first consider the problem of concurrent sensory and motor adaptation. Previous models have assumed that sensory and motor adaptation occur independently from one another, considering one model for sensory adaptation (e.g., Ghahramani, Wolpert, & Jordan, 1997), and another for motor adaptation (e.g., Donchin, Francis, & Shadmehr, 2003). We show that, by considering a full generative model of the joint observations and the disturbances that affect them, a unified model of sensory and motor adaptation can be derived that makes strong and experimentally verifiable predictions about interactions between sensory and motor adaptation (Haith, Jackson, Miall, & Vijayakumar, 2008). Next, we consider the problem of multisensory oddity detection. Common naive normative models of cue combination are not robust, falsely predicting the existence of infinitely many discrepant but still indistinguishable stimuli. A full generative model of the process is required to explain this entire domain of human behavior (Hospedales & Vijayakumar, 2009).

INTERACTIONS BETWEEN SENSORY AND MOTOR ADAPTATION

Many chapters in this book focus on problems associated with combining multiple, possibly discrepant cues. If, however, two cues are persistently discrepant by the same amount, it is likely that there is a systematic miscalibration of one modality or the other. For example, prism goggles can be worn which shift the entire visual field, introducing a discrepancy between visual and proprioceptive estimates of hand position. Such discrepancies can be eliminated by adapting the senses over time so that they become realigned.

Previous Models of Sensory Adaptation

If the hand is viewed through prism goggles, a realignment takes place between vision and proprioception with, typically, a shift in the visual estimate of hand position and an opposite shift in the proprioceptive estimate of hand position (Redding & Wallace, 1996). We model sensory adaptation by assuming that the visually and proprioceptively observed hand positions are displaced by some systematic disturbances (i.e., miscalibrations or unknown experimental manipulations), with added Gaussian noise

$$v_t = y_t + r_t^v + \varepsilon_t^v, \tag{4.1}$$

$$p_t = y_t + r_t^p + \varepsilon_t^p. \tag{4.2}$$

Here v_t and p_t are the subject's visual and proprioceptive observations of their hand position. r_t^v and r_t^p are miscalibrations of vision and proprioception, and ε_t^v and ε_t^p represent observation noise corrupting each measurement, which we assume to be Gaussian with variance σ_v^2 and σ_p^2, respectively. We assume that the subject maintains estimates $\hat{r}_t^v$ and $\hat{r}_t^p$ of each disturbance over time. The subject's visual and proprioceptive estimates of hand position will be given by subtracting the relevant disturbance estimates from their observations, that is,

$$\hat{y}_t^v = p_t - \hat{r}_t^v, \tag{4.3}$$

$$\hat{y}_t^p = p_t - \hat{r}_t^p. \tag{4.4}$$

The maximum likelihood estimate (MLE) of the true hand position y_t is given by

$$\hat{y}_t^{MLE} = \frac{\sigma_p^2}{\sigma_v^2 + \sigma_p^2}\hat{y}_t^v + \frac{\sigma_v^2}{\sigma_v^2 + \sigma_p^2}\hat{y}_t^p. \quad (4.5)$$

This estimate optimally combines the two unimodal estimates into a single estimate, taking into account the relative observation noise in each modality. Ghahramani et al. (1997) proposed that we adapt the estimates of $\hat{r}_t^v$ and $\hat{r}_t^p$ in such a way that the maximum likelihood estimate (MLE) of the actual hand position remains unchanged, which leads to the following update for the disturbance estimates:

$$\hat{r}_{t+1}^v = \hat{r}_t^v + \eta w_p[\hat{y}_t^p - \hat{y}_t^v], \quad (4.6)$$

$$\hat{r}_{t+1}^p = \hat{r}_t^p + \eta w_v[\hat{y}_t^v - \hat{y}_t^p], \quad (4.7)$$

where η is some fixed adaptation rate and w_p and w_v are the respective combination weights in Eq. 4.5. From a statistical learning viewpoint, this model can be understood as treating the miscalibrations $\hat{r}_t^v$ and $\hat{r}_t^p$ as unknown *parameters*, which are estimated via an online variant of the standard expectation-maximization algorithm (Bishop, 2006) for parameter estimation in statistical models. The corresponding graphical model is illustrated in Figure 4.1. A crucial prediction of this model is that sensory adaptation will be driven purely by discrepancy between the two senses. This model can successfully account for many features of sensory adaptation, particularly in purely passive contexts such as visual-auditory integration; it has also been proposed as a model for adaptation in visual-proprioceptive integration during active movement (van Beers, Wolpert, & Haggard, 2002).

This model, however, is not quite sufficient on its own to explain adaptation of reaching movements during exposure to shifts in visual feedback. While a recalibration of the visual system will be reflected in reaches toward visual targets, the extent of visual adaptation is always less than the experimentally imposed visual shift. The fact that subjects can nevertheless reach the target successfully implies that they additionally learn a correction to their movements as well as compensating their perceptual estimates of hand and target locations. Simani, McGuire, and Sabes (2007) recently demonstrated that the task performed during exposure affects generalization to reach trials after the visual shift is removed. This difference would not occur if the adaptation were purely sensory in nature.

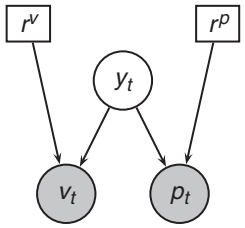

Figure 4.1 Graphical model for a MLE-based sensory-adaptation model. Shaded circles represent observed random variables. Unshaded circles represent unobserved random variables. Squares represent unknown parameters. Noisy visual and proprioceptive observations, v_t and p_t of unknown hand position y_t are available at each trial/time step. These may be subject to unknown biases r^v and r^p due to miscalibration or experimental manipulation. In the MLE-based model, these unknown biases are treated as parameters of the model which are estimated via online expectation maximization.

Although no explicit model of concurrent sensory and motor adaptation has been previously proposed, it is straightforward to augment the aforementioned sensory-adaptation model with a standard state-space model of motor adaptation. We assume that a motor disturbance affects the relationship between the subject's motor commands and the position of the hand at the end of the movement. Specifically

$$y_t = u_t + r_t^y + \varepsilon_t^y, \quad (4.8)$$

where u_t is the subject's motor command, r_t^y is the motor disturbance acting on the hand, and $\varepsilon_t^y \sim N(0, \sigma_y^2)$ is motor execution noise. Existing state-space models of motor adaptation (e.g., Donchin et al., 2003) typically assume that an estimate of this disturbance is updated according to the error in the hand position midway through

the movement. Although the subject does not know the true error in the hand position (since only noisy, corrupted observations of hand position are available), the hand position error can be estimated using the hand position MLE, leading to the following learning rule

$$\hat{r}^y_{t+1} = \hat{r}^y_t + \xi[\hat{y}^*_t - \hat{y}^{MLE}_t], \tag{4.9}$$

where v^* is the visually observed target position, and $\hat{y}^*_t = (v^* - \hat{r}^v_t)$ is the estimated desired hand location, and ξ is some fixed adaptation rate.

This combined model reflects the view that sensory and motor adaptations are distinct processes. The sensory-adaptation component is driven purely by discrepancy between the senses, while the motor-adaptation component only has access to a single, fused estimate of hand position and is driven purely by estimated performance error.

Bayesian Sensory- and Motor-Adaptation Model

We propose an alternative approach to solving the sensorimotor-adaptation problem. Rather than modeling sensory and motor adaptation independently, we consider a full generative model of how sensory and motor disturbances affect a subject's visual and proprioceptive observations. All three disturbances are now treated as random variables that the subject is attempting to estimate simultaneously. This model is illustrated in Figure 4.2.

The subject generates a motor command u_t, which leads to a new hand position y_t, perturbed by some unknown motor disturbance r^y_t as well as motor noise ε^y_t, as in Eq. 4.8. This hand position is not directly observed, but noisy and potentially biased visual and proprioceptive observations are available, as described in Eqs. 4.1 and 4.2.

In addition to this statistical model of how actions and observations are affected by the three disturbances, r^v_t, r^y_t, and r^p_t, the subject has some beliefs about how these disturbances evolve over time. These beliefs are characterized by a trial-to-trial disturbance dynamics model given by

$$\mathbf{r}_{t+1} = A\mathbf{r}_t + \eta_t, \tag{4.10}$$

where A is some diagonal matrix and η_t, is a random drift term with zero mean and diagonal

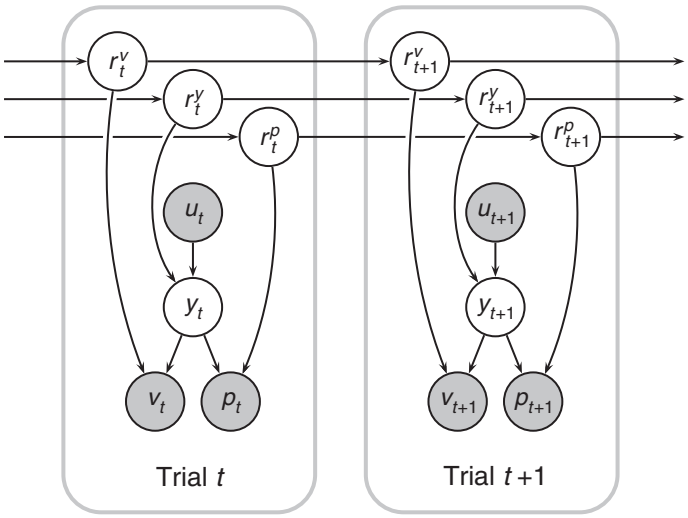

Figure 4.2 Bayesian sensory- and motor-adaptation model. Shaded circles represent observed random variables (motor command u_t, visual and proprioceptive observations v_t and p_t). Unshaded circles represent unobserved random variables (hand position y_t, visual and proprioceptive miscalibrations r^v_t and r^p_t and motor disturbance r^y_t).

covariance matrix Q, that is,

$$\eta_t \sim N(0, Q). \tag{4.11}$$

A and Q are both diagonal to reflect the fact that each disturbance evolves independently. We denote the diagonal elements of A by $\mathbf{a} = (a^v, a^p, a^y)$ and the diagonal of Q by $\mathbf{q} = (q^v, q^p, q^y)$. The vector $\mathbf{a}$ describes the timescales over which each disturbance persists, whereas $\mathbf{q}$ describes the random drift in the disturbance from one trial to the next. These parameters reflect the statistics of the usual fluctuations in sensory calibration errors and motor plant dynamics, which the sensorimotor system must adapt to on an ongoing basis. (Similar assumptions have previously been made elsewhere [Körding, Tenenbaum, & Shadmehr, 2007; Krakauer, Mazzoni, Ghazizadeh, Ravindran, & Shadmehr, 2006]).

We propose that the patterns of adaptation and the sensory after-effects exhibited by subjects correspond to optimal inference of the disturbances $\mathbf{r}_t$ within this full generative model, given the observations on each trial. This is in contrast to alternative models presented earlier in which sensory and motor adaptation are assumed to be mediated by independent processes.

The linear dynamics and Gaussian noise of the observer's model mean that the posterior probability of the disturbances given the observations can be calculated analytically, and it becomes equivalent to a Kalman filter. The latent state tracked by the Kalman filter is the vector of disturbances $\mathbf{r}_t = (r_t^v, r_t^p, r_t^y)^T$, with state dynamics given by Eq. 4.10. The observations v_t and p_t are related to the disturbances via

$$\begin{pmatrix} v_t \\ p_t \end{pmatrix} = \begin{pmatrix} u_t \\ u_t \end{pmatrix} + \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} (\mathbf{r}_t + \varepsilon_t), \tag{4.12}$$

where $\varepsilon_t = (\varepsilon_t^v, \varepsilon_t^p, \varepsilon_t^y)^T$. We can write this in a more conventional form as

$$\mathbf{z}_t = H\mathbf{r}_t + H\varepsilon_t \tag{4.13}$$

where $\mathbf{z}_t = (v_t - u_t,\ p_t - u_t)^T$ and $H = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$. The observation noise covariance is given by

$$R = E\left[(H\varepsilon_t)(H\varepsilon_t)^T\right] = \begin{pmatrix} \sigma_v^2 + \sigma_y^2 & \sigma_y^2 \\ \sigma_y^2 & \sigma_p^2 + \sigma_y^2 \end{pmatrix}, \tag{4.14}$$

where σ_y^2 is motor execution noise, and σ_v^2 and σ_p^2 represent the noise in the subjects visual and proprioceptive estimates as before. The standard Kalman filter update equations can be used to predict how a subject will update estimates of the disturbances following each trial and therefore what actions to select on the next trial, leading to a full prediction of performance from the first trial onward.

Experiment: Testing Sensory Adaptation during Force-Field Exposure

While the MLE-based model predicts there will be sensory adaptation *only* when there is a discrepancy between the senses, the Bayesian model predicts that there will also be sensory adaptation in response to a motor disturbance (such as an external force applied to the hand). Just as a purely visual disturbance can lead to a multifaceted adaptive response, the Bayesian models predicts that a purely motor disturbance will result in both motor and sensory adaptation, even though there is never any discrepancy between the senses. This occurs because there are three unknown disturbances, but only two observation modalities on each trial. There are therefore many combinations of disturbances that can account for the observations on each trial. Because of the subject's assumptions about how the disturbances vary over time (i.e., Eq. 4.10), explanations that assign credit to all three disturbances are more likely than the true disturbance that was experienced.

We experimentally tested the hypothesis that force-field adaptation would lead to sensory adaptation. We tested 11 subjects who performed a series of trials consisting of reaching movements interleaved with perceptual-alignment tests.

Subjects grasped the handle of a robotic manipulandum with their right hand. The hand was not visible directly, but a cursor displayed via a mirror/flat-screen-monitor setup (Fig. 4.3A) was exactly coplanar and aligned with the handle of the manipulandum. In the movement phase, subjects made an out-and-back reaching movement toward a visual target with their right hand. In the visual localization phase, a visual target was displayed pseudorandomly in one of five positions and the subjects moved their left fingertip to the perceived location of the target. In the proprioceptive localization phase, the right hand was passively moved to a random target location, with no visual cue of its position, and subjects moved their left fingertip to the perceived location of the right hand. Left fingertip positions were recorded using a Polhemus motion tracker. Neither hand was directly visible at any time during the experiment.

Subjects were given 25 baseline trials with zero external force, after which a force field was gradually introduced (Fig. 4.3B). A leftward lateral force F_x was applied to the right hand during the reaching phase. The magnitude of the force was proportional to the forward velocity $\dot{y}$ of the hand, that is,

$$F_x = -a\dot{y}. \tag{4.15}$$

The force was applied only on the outward part of the movement (i.e., only when $\dot{y} > 0$). After steadily incrementing a during 50 adaptation trials, the force field was then kept constant at $a = 0.3N/(cm\ s^{-1})$ for a further 25 postadaptation test trials.

We compared the average performance in the visual and proprioceptive alignment tests before and after adaptation in the velocity-dependent force field. The results are summarized in Figure 4.4A. Most subjects exhibited small but significant shifts in alignment bias in both the visual- and proprioceptive-alignment tests. Two subjects exhibited shifts that were more than two standard deviations away from the average shift and were excluded from the analysis. We found significant lateral shifts in both visual and proprioceptive alignment bias in the direction of the perturbation (both $p < .05$, one-tailed paired t-test). In the y-direction, the initial alignment bias was very high. However, there was no significant shift in either modality (Fig. 4.4B),

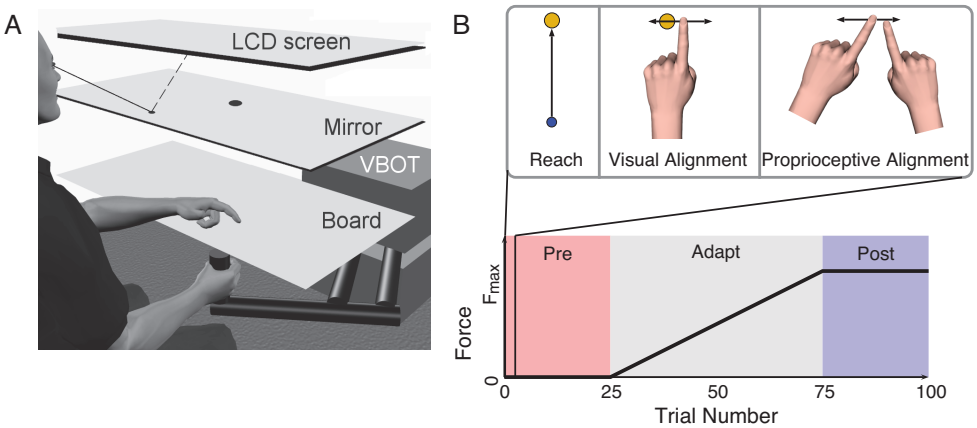

Figure 4.3 (A) Experimental setup. (B) Subjects made reaching movements in a single direction while perturbed by a force field, the magnitude of which was gradually increased over trials. These reaching movements were interleaved with perceptual alignment tests to measure the extent of sensory recalibration. In these alignment tests, subjects moved their (unseen) left hand to align it as best as possible with either a visual target, or their (unseen) right hand.

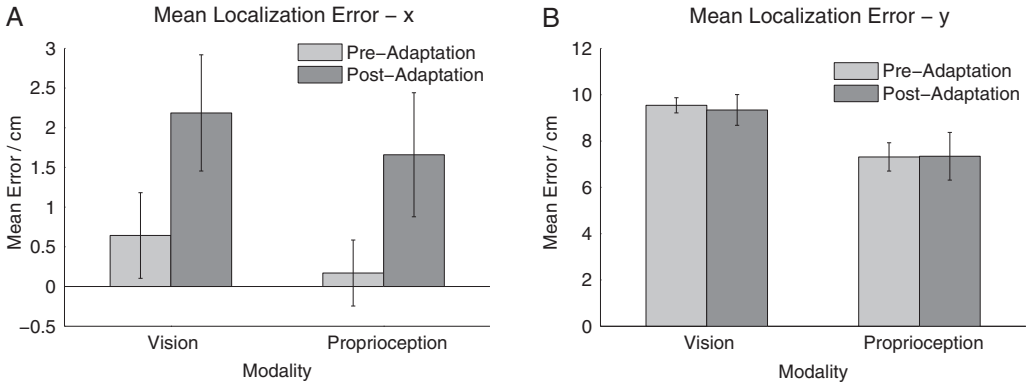

Figure 4.4 Comparison of visual and proprioceptive alignment biases before vs. after adaptation, (A) in the direction of and (B) perpendicular to the perturbation.

consistent with the fact that there was no perturbation in this direction.

We assessed subjects' ability to counteract the force during the reach trials by measuring the amount by which subjects missed the target. We quantified this as the perpendicular distance between the furthest point in the trajectory and the straight line passing through both the start position and the target. We fitted the Bayesian and MLE-based models to the data using nonlinear optimization to minimize the squared error between the model and the data across the alignment tests and reach performance. Figure 4.5 shows the averaged data along with the model fits. Both models were able to account similarly well for the trends in reaching performance across trials (Fig. 4.5A). Figures 4.5B and 4.5C show the model fits for the alignment tasks. The Bayesian model is able to account for both the extent of the shift and the time course of this shift during adaptation. Since there was never any sensory discrepancy, the MLE-based model predicted no change in the localization task.

These results support the prediction of the Bayesian model that adaptation to a force field would also lead to sensory adaptation. Alternative models in which sensory and motor adaptation are considered to be independent processes fail to predict this effect, since there is never any discrepancy between the two sensory estimates of hand position. Furthermore, the Bayesian model was able to account accurately for the trends in both reaching performance and alignment-test errors on a trial-to-trial basis, strongly suggesting that the brain uses the principles of Bayesian estimation to guide adaptation.

The brain in this case can be considered to act as an "ideal observer," since it makes the best possible use of all information that it receives through application of an appropriate generative model capturing the dependence of its observations on unknown features of the environment. In the next section, we show how this same general principle can be applied in a different context where information from visual and haptic modalities must be combined to guide decision making in an oddity detection task.

MULTISENSORY ODDITY DETECTION AS BAYESIAN INFERENCE

Bayesian ideal-observer modeling has been applied extensively and successfully to understand tasks that require integration of two or more cues in the estimation of some real-world stimulus. Much of this work makes common, but simple generative modeling assumptions of independence with Gaussian noise, under which the ideal-observer's strategy for inference in these generative models has the particularly

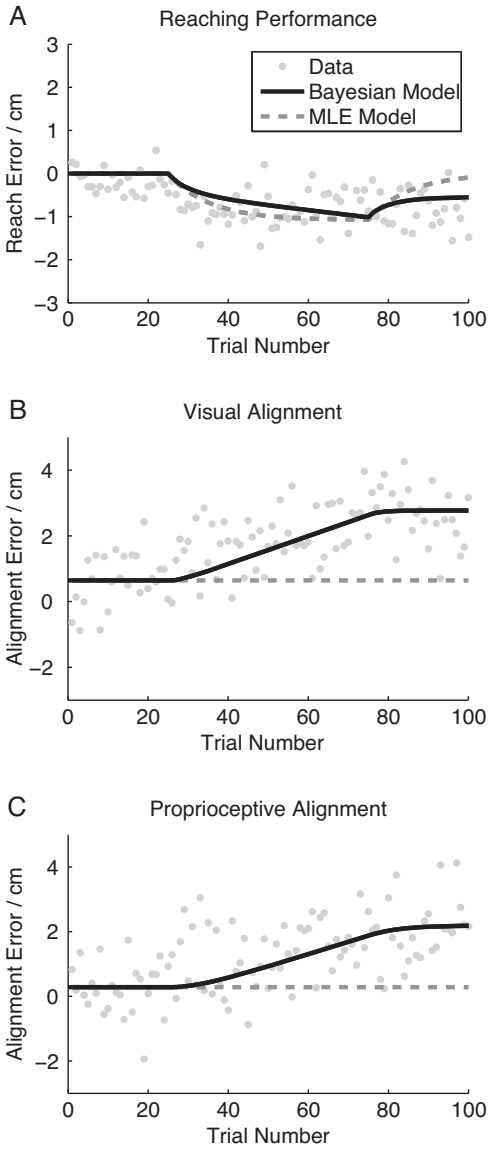

Figure 4.5 Trial-by-trial reaching and alignment test performance, with model fits. (A) Reaching error. (B) Alignment bias during left-hand-to-visual alignment tests, corresponding to $\hat{r}_t^v$ in the models. (C) Alignment bias during right hand to visual alignment tests, corresponding to $\hat{r}_t^p$ in the models.

simple form of reliability-weighted linear cue combination (see Chapter 1). We will refer to this as maximum likelihood integration (MLI). This approach presumes that the *correspondence* between observations and latent variables (relevant unknown aspects of world state) is obvious and therefore unnecessary to model.

In some cases, experimenters have relaxed this assumption and provided subjects with stimuli where this correspondence (causal structure) was not obvious. That is, it was not obvious which of multiple possible sources caused the observations (Hairston et al., 2003; Shams, Kamitani, & Shimojo, 2000; Shams, Ma, & Beierholm, 2005; Wallace et al., 2004), or which of multiple possible world models was true (Knill, 2007). In these cases, the standard MLI linear-cue-combination approach fails to explain human performance. As we shall see, this seems not to be due to suboptimality of human perception, but mismatch between the experiments and overly simple experimental models.

Under the generative modeling approach proposed in this chapter, we see that uncertain correspondence in a perceptual problem corresponds to uncertain *structure* in the generative model. An ideal Bayesian observer should also infer this uncertain structure. Recently, studies have begun to apply a complete Bayesian-structure-inference perspective (Hospedales & Vijayakumar, 2008) to experiments with correspondence or structure uncertainty and have provided a good explanation for the human perception in these cases (Chapters 2 and 13; Körding, Beierholm et al., 2007).

Here, we consider the challenging modeling problem of *multisensory oddity detection*, in which we shall see that structure uncertainty occurs simultaneously in two different ways. We show how to formalize a generative model of this problem, and how this can explain and unify a pair of experiments (Hillis, Ernst, Banks, & Landy, 2002) where MLI previously failed dramatically.

Next, we briefly review standard MLI ideal-observer modeling for cue combination, and show—by way of theoretical argument as well as a concrete experimental example—why the

naive application of mandatory MLI approaches qualitatively fails to explain human multisensory oddity detection.

Standard Ideal-Observer Modeling for Sensor Fusion

A generative probabilistic model (Bishop, 2006) for a perceptual problem describes the way in which signals are generated by a source, and how they are then observed—including any distorting noise processes. Predictions made by the results of optimal inference in this model can then be compared to experimental results.

Formalized as a generative model, standard cue-combination theory (Fig. 4.6) assumes that multisensory observations x_m in modalities m are generated from some source y in the world, subject to independent noise in the environment and physical sensor apparatus, for example, $x_m \sim N(y, \sigma_m^2)$. Ernst and Banks (2002) asked subjects to make haptic x_h and visual x_v observations of a bar's height y and estimate the true height ($\hat{y}_{h,y}$) in order to compare the sizes of two bars. This requires computing the posterior distribution over height, which under these modeling assumptions is Gaussian $p(y|x_h, x_v; \sigma_h^2, \sigma_v^2) = N(y; \mu_{y|h,v}, \sigma_{y|h,v}^2)$, with mean and variance given by Eqs. 4.16–4.17:

$$\mu_{y|h,v} = \frac{\sigma_h^{-2}}{\sigma_h^{-2} + \sigma_v^{-2}} x_h + \frac{\sigma_v^{-2}}{\sigma_h^{-2} + \sigma_v^{-2}} x_v, \tag{4.16}$$

$$\sigma_{y|h,v}^2 = \frac{\sigma_h^2 \sigma_v^2}{\sigma_h^2 + \sigma_v^2}. \tag{4.17}$$

Psychophysics experiments (e.g., Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003) typically test multisensory perception for optimality by matching to the ideal-observer performance in two ways. First, the optimal estimate of the true height is $\hat{y}_{h,v} = \mu_{y|h,v}$, so the variance of the human's responses $\hat{y}_{h,v}$ to multisensory stimuli should match the variance of the optimal response $\sigma_{y|h,v}^2$ (Eq. 4.17). Note from Eq. 4.17 that this is always less than the variance of individual observations σ_h^2, σ_v^2; hence, it is less than the variance of the unimodal responses $\hat{y}_h$, $\hat{y}_v$. Secondly, the multisensory response of the ideal observer is the precision-weighted mean of the unimodal observations (Eq. 4.16). Therefore, experimentally manipulating the variances σ_h^2, σ_v^2 of the individual modalities should produce the appropriate changes in the human perceptual response $\hat{y}_{h,v}$. These quantities can be determined directly in direct-estimation experiments (e.g., Wallace et al., 2004) or indirectly via fitting a psychometric function in two-alternative forced-choice experiments (e.g., Alais & Burr, 2004; Battaglia et al., 2003).

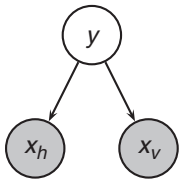

Figure 4.6 Standard sensor-fusion model. Bar size y is inferred on the basis of haptic and visual observations x_h and x_v (Hillis et al., 2002). Shaded circles indicate observed quantities and empty circles indicate quantities to estimate.

Oddity Detection

In direct-estimation scenarios, subjects try to make a *continuous estimate* of a particular unknown quantity y, such as the height of a bar or spatial stimulus location based on noisy observations x_m, such as visual and haptic heights or auditory and visual locations, respectively. In contrast, in the *oddity-detection paradigm*, subjects observe $i = 1 \ldots N$ separate stimuli $x_{m,i} \sim N(y_i, \sigma_m^2)$ and must make a discrete determination of the "odd" stimulus o from among the $N \geq 3$ options $\{y_i\}_{i=1}^N$. Depending on the experimental paradigm, the odd stimulus may be detectable because it is, for example, larger or smaller than the other stimuli.

Multisensory oddity detection is a particularly interesting problem to study because it provides novel paradigms for manipulating the oddity. Specifically, the *mandatory* MLI theory of cue combination predicts that a single fused estimate $\hat{y}_{h,v}$ will be made for each multisensory stimulus (Eq. 4.16), and oddity detection will proceed solely based on these estimates. This means

that a particular stimulus might be the same as the others when averaged over its modalities of perception (Eq. 4.16), while each individual stimulus modality could simultaneously be radically discrepant. Such stimuli would be known as *perceptual metamers*, meaning that although they would be physically distinct, they would be perceptually indistinguishable under this theory of cue combination. This provides a new and interesting test of Bayesian perception, because if the nervous system was to use solely the fused estimates to detect oddity, then it would not be able to discriminate such metamers. If, on the other hand, the nervous system made an inference about structure in the full generative model of the observations, it could detect such stimuli on the basis of structure (correspondence) oddity. In the following section, we formalize this inference paradigm and look in detail at a pair of experiments that tested oddity detection and found MLI mandatory fusion models unsatisfactory in explaining the data completely.

Human Multisensory Oddity Detection Performance

Hillis et al. (2002) studied multisensory oddity detection in humans using $N = 3$ options in two conditions: visual-haptic cues for size (across-modal cues) and texture-disparity cues for slant (within-modal cues). We describe this experiment in some detail and will formalize the oddity-detection problem and our solution to it in the context of this experiment. It should be noted that our approach can trivially be generalized to other conditions, such as more modalities of observation and selecting among $N \geq 3$ options.

Three stimuli are presented in two modalities v and h (Fig. 4.7). (To simplify the discussion, we will refer generally to the visual-haptic (v-h) modalities when discussing concepts which apply to both the visual-haptic and texture-disparity experiments.) Two of the stimuli are instances of a fixed standard stimulus y_s and one is an instance of the (potentially odd) probe stimulus y_o. The standard stimuli are always *concordant*, meaning that there is no experimental manipulation across modalities so $y_s = y_{h,s} = y_{v,s}$. The probe stimulus y_o is experimentally manipulated across a wide range of values so that the visual and haptic sources, $y_{v,o}$ and $y_{h,o}$, may or may not be similar to each other or to the standard y_s. The subject's task is to detect which of the three stimuli is the probe. If all the stimuli are concordant and the probe is set the same as the standard $y_s = y_o$, then we expect no better than random (33%) success rate (Fig. 4.7A). If all the stimuli are concordant and the probe discrepancy is set very high compared to the standard, then we expect close to 100% success rate (Fig. 4.7B). However, if the probe stimulus is experimentally manipulated to be discordant so that $y_{h,o} \neq y_{v,o}$, then the success rate expected will depend on precisely how the subjects combine their observations of $y_{h,o}$ and $y_{v,o}$ (Fig. 4.7C). The two-dimensional distribution of detection success/error rate as a function of controlled probe values $y_{h,o}$, $y_{v,o}$ can be measured and used to test different theories of cue combination.

For a single modality, for example, h, the error rate distribution for detection of the probe $y_{h,o}$ can be approximated as a one-dimensional Gaussian bump centered at the standard $y_{h,s}$. (If $y_{h,s} = y_{h,o}$ then detection of the odd stimulus will be at chance level; if $y_{h,o} \gg y_{v,o}$ then detection of the odd stimulus will be reliable, etc.) The shape of the two-dimensional performance surface for multimodal probe stimulus detection $p(\text{success}|y_{h,o}, y_{v,o})$ can be modeled as a two-dimensional bump centered at (y_s, y_s). Performance *thresholds* (the equipotentials where $p(\text{success}|y_{h,o}, y_{v,o}) = 66\%$) are computed from the performance surfaces predicted by theory and those of the experimental data. Cue-combination theories are evaluated by the match of their predicted thresholds to the empirical thresholds.

Basic Cue-Combination Theories

To parameterize models for testing, the observation precisions first need to be determined. Following standard practice for MLI modeling, Hillis et al. (2002) measured the variances of the unimodal error distributions and then used these to predict the combined variance and hence the

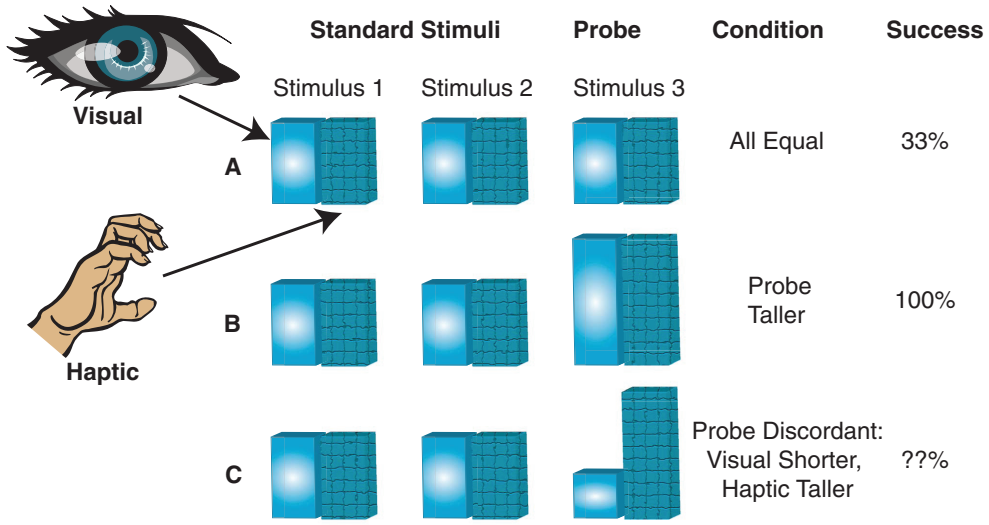

Figure 4.7 Schematic of visual-haptic height oddity-detection experimental task from (Hillis et al., 2002). Subjects must choose the odd probe stimulus based on haptic (textured bars) and visual (plain bars) observation modalities. (A) Probe stimulus is the same as the standard stimuli: detection at chance level. (B) Probe stimulus bigger than standard: detection is reliable. (C) Haptic and visual probe modalities are discordant: detection rate will depend on cue-combination strategy.

multimodal error distribution under MLI cue-combination theory (Eqs. 4.16–4.17). (In the next section, we will discuss why this approach is not quite ideal for this experiment.)

Specifically, under the MLI theory, the brain would compute a fused estimate $\hat{y}_o$ based on the two observations $x_{h,o}, x_{v,o}$ (Eqs. 4.16–4.17) and then discriminate based on this estimate. In this case, although both cues are now being used, some combinations of cues would produce a metameric probe, that is, physically distinct but perceptually indistinguishable. Specifically, if we parameterize the probe stimuli as $y_{h,o} = y_{h,s} + \Delta y_{h,o}$, $y_{v,o} = y_{v,s} + \Delta y_{v,o}$, then along the diagonal line through the performance surface where $\Delta y_{h,o} = -(\sigma_v^2/\sigma_h^2)\Delta y_{v,o}$, the fused estimate is on average the same as the standard $\hat{y}_{h,o} = y_s$ and the probe would be undetectable. Performance along the cues-concordant diagonal, however, would be improved compared to the single-cue estimation cases because the combined variance is less than the individual variances ($\sigma^2_{y|h,v} < \sigma_h^2$ and $\sigma^2_{y|h,v} < \sigma_v^2$).

Two variants of the experiment were performed, one for size discrimination across visual and haptic modalities (standard: $y_s = 55$ mm), and one for slant discrimination using texture and stereo disparity cues within vision (standard: $y_s = 0$ deg). Figure 4.8 illustrates the predicted performance surface contours for unimodal models (red lines), the MLI model (green lines), and those observed (dots) by Hillis et al. (2002) for two sample subjects. Contour points closer to the origin y_s indicate better performance.

There are several points to note in Figure 4.8: (1) In the cues-concordant quadrants (1 and 3), the multimodal performance is improved compared to the unimodal performance, as predicted by the MLI theory (magenta points and green lines are inside the red lines in quadrants 1 and 3). (2) Particularly in the intramodal case (Fig. 4.8B), the observed experimental performance is significantly worse than the unimodal performance in the cues-discordant quadrants (2 and 4) (magenta points are outside of the red lines in Fig. 4.8B, quadrants 2 and 4). Note that the green intramodal predicted thresholds in Figure 4.8B are curved, unlike the straight intermodal predicted thresholds in

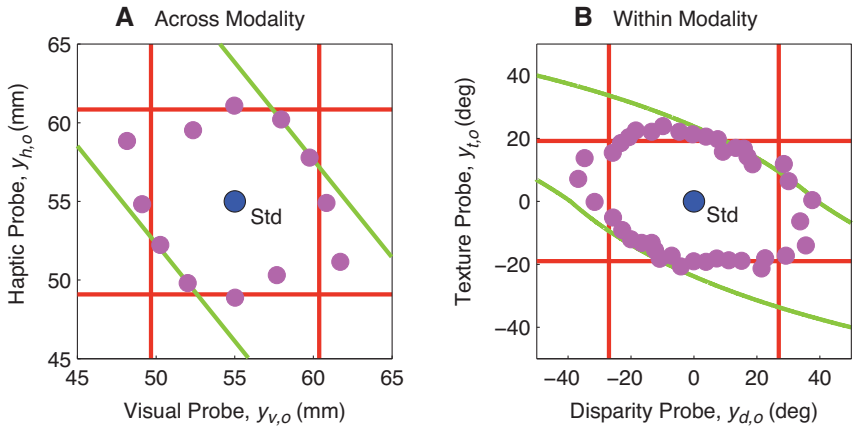

Figure 4.8 MLI oddity-detection predictions and experimental results. (A) Visual-haptic experiment. (B) Texture-disparity experiment. Red lines: Observed uni-modal discrimination thresholds. Green lines: Discrimination-threshold predictions assuming mandatory fusion. Magenta points: Discrimination threshold observed experimentally for two sample subjects from Hillis et al. (2002).

Figure 4.8A. This is due to the use of a slightly more complicated model than described here, which reflects the fact that the variance of the slant cue σ_s^2 itself depends on the current slant y_s (see Hillis et al., 2002, for details). The essential insights remain the same, however.

Hillis et al. concluded that mandatory fusion applied within (Fig. 4.8B) but not between (Fig. 4.8A) the senses, in part because poor performance in the cues-discordant quadrants 2 and 4 was noted to be less prominent in the intermodal case. They hypothesized that the discrepancy between the observed limited region of poor performance in the intramodal cues-discordant quadrants 2 and 4, and the MLI predicted infinite region of nondiscriminability could be due to a separate texture consistency mechanism ultimately enabling the discrimination in quadrants 2 and 4 (Hillis et al., 2002).

Nevertheless, the classical unifying theory of ideal-observer maximum-likelihood combination retains a strong *qualitative* discrepancy with the experimental results (Fig. 4.8, green lines and points) in both experiments. It does not predict good performance in the cues-concordant quadrants 1 and 2 as well as a *limited* region of poor performance in the cues-discordant quadrants 2 and 4. In the next sections, we will show how an alternative unifying approach, exploiting a complete generative model of the oddity-detection problem, including the associated structure uncertainty, can explain both of these experiments quantitatively and intuitively.

Modeling Oddity Detection

The classical MLI approach to sensor fusion has failed as a means to understand human performance in this multisensory oddity-detection problem. Let us step back and reconsider the match between the problem and its generative model. There are two key components of this problem that are not modeled by the classical approach (Fig. 4.6): the discrete model-selection nature of the problem, and the variable structure component of the problem.

The task posed—"*Is stimulus 1, 2 or 3 the odd one out?*"—is actually no longer simply an estimation of a combined stimulus $\hat{y}_{h,v}$. This estimation is involved in solving the task, but ultimately the task effectively asks subjects to make a *probabilistic model selection* (Mackay, 2003) between three models. To understand the model-selection interpretation intuitively, consider the following reasoning process: I have experienced three noisy multisensory observations. I do not know the true values of these three stimuli, but I know they come from

two categories, standard and probe. Which of the following is more plausible:

1. Multisensory stimuli two and three come from one category, and stimulus one comes from another.
2. Stimuli one and three come from one category, and stimulus two comes from a different category.
3. Stimuli one and two come from one category, and stimulus three comes from another.

With this in mind, to take a Bayesian ideal-observer point of view on this experiment, the experimental task is clearly to estimate which of three distinct *models* is the best one for the data. That is, the experiment effectively asks which model in an entire set of models best explains the data, rather than asking the value of some variable within a model. The ideal-observer should *integrate over the distribution of unknown stimulus values* y_s and y_o (since subjects are not directly asked about these) in determining the most plausible model (assignment of oddity).

The second key aspect of this task which must be included in any full generative model of this problem is that oddity can be entailed in the probe stimulus not only by its combined difference from the standard, but by discrepancy within the probe stimulus. In this case (similar to other recent multisensory perception experiments with variable causal structure: Hairston et al., 2003; Shams et al., 2000, 2005; Wallace et al., 2004), the variable structure can effectively "give away" the probe. We introduced the approach needed to solve this type of problem in multisensory perception as *structure inference* (Hospedales & Vijayakumar, 2008). Körding, Beierholm et al. (2007) carried out a detailed analysis of the experiments of Hairston et al. (2003) and Wallace et al. (2004) and showed how the structure inference approach was necessary to explain the results, but they termed the procedure *causal inference* (see Chapter 2).

Formalizing Optimal Oddity Detection

A generative-model Bayesian network formalization of the oddity-detection task for the three multisensory observations $\{x_{h,i}, x_{v,i}\}_{i=1}^{3}$ is shown in Figure 4.9, where the aim is to determine which observation is the odd probe. The graph on the left indicates that the four observations composing the other two standard stimuli are all related to the standard stimulus value y_s. The graph on the right indicates that the probe visual-haptic observations are independent of the standard but *might* be related via their common parent, the latent probe stimulus of value y_o. The latent variable C switches whether the probe observations have a common cause in the generative model. The prior probability of common causation is given

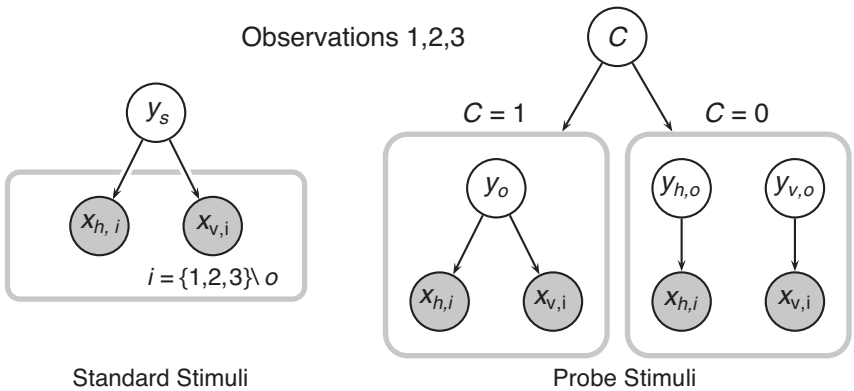

Figure 4.9 Graphical model for oddity detection via structure inference. A subject observes three multisensory stimuli, $x_{h,i}$ and $x_{v,i}$, $i = 1,2,3$. The three options for assigning oddity correspond to three possible models indexed by $o = 1,2,3$. The uncertain causal structure of the probe stimulus is now represented by C, which is computed in the process of evaluating the likelihood of each model o.

by the new parameter π_c, so $p(C=1)=\pi_c$ and $p(C=0)=1-\pi_c$. Under the hypotheses of common causal structure $C=1$, we assume that the two observations $x_{h,o}, x_{v,o}$ were produced from a single latent variable y_s. Alternately, if $C=0$, we assume separate sources $y_{h,o}$ and $y_{v,o}$ were responsible for each. An ideal Bayesian observer in this task should integrate over both the unknown stimulus values and the causal structure C (i.e., whether we are feeling and seeing the same thing).

The three different possible models are given by the different probe hypotheses $o = 1, 2, 3$, which separate the standard and probe stimuli into different clusters. We represent this clustering in terms of the *set difference operator* "\." For example, $o=3$ would mean that stimuli $\{1,2,3\}\backslash 3 = \{1,2\}$ are drawn from the standard y_s, and therefore observations $\{x_{h,1}, x_{v,1}, x_{h,2}, x_{v,2}\}$ (Fig. 4.9, left) should be similar to each other—and potentially dissimilar to odd probe observations $\{x_{h,3}, x_{v,3}\}$ (Fig. 4.9, right), which were generated independently. The ideal Bayesian observer would base its estimation of oddity on the marginal likelihood of each stimulus/model o being odd, $p(o|\{x_{h,i}, x_{v,i}\}_{i=1}^3\theta) \propto p(\{x_{h,i}, x_{v,i}\}_{i=1}^3|o,\theta)p(o|\theta)$:

$$
\begin{aligned}
&p(\{x_{h,i},x_{v,i}\}_{i=1}^3|o,\theta)\\
&\quad = p_s(\{x_{h,i},x_{v,i}\}_{i\in\{1,2,3\}\backslash o}|o,\theta)p_o(x_{h,o},x_{v,o}|o,\theta),\\
&p_s(\{x_{h,i},x_{v,i}\}_{i\in\{1,2,3\}\backslash o}|o,\theta)\\
&\quad = \int \prod_{i\in\{1,2,3\}\backslash o}\prod_{j=h,v} N(x_{j,i}|y_s,\theta)N(y_s|\theta)dy_s,\\
&p_o(x_{h,o},x_{v,o}|o,\theta) = \int \prod_{j=h,v} N(x_{j,o}|y_o,\theta)N(y_o|\theta)\\
&\quad \times \pi_c dy_o + \iint \prod_{j=h,v} N(x_{j,o}|y_{j,o},\theta)N(y_{j,o}|\theta)\\
&\quad \times (1-\pi_c)dy_{h,o}dy_{v,o}. \qquad (4.18)
\end{aligned}
$$

The marginal likelihood factors into a product of standard p_s and probe p_o parts, which may be decomposed into integrals of Gaussian products, which are simple to evaluate analytically (see Hospedales & Vijayakumar, 2009, for more details). This procedure evaluates how likely each stimulus o is to be odd, accounting for the uncertainty in stimulus values y (integrals) and the uncertain causality of the probe data C (sum). Here, θ summarizes all the fixed model parameters, for example, the observation variances σ_h^2 and σ_v^2.

Results

To evaluate our multisensory oddity-detection model, we assume no prior preference for which stimulus is odd ($p(o)$ uniform) and therefore estimate the probe based on the likelihood, $\hat{o} = \arg\max_o p(\{x_{h,i}, x_{v,i}\}_{i=1}^3|o,\theta)$. Evaluating the detection success rate for a range of probe values $y_{v,o}$ and $y_{h,o}$, we can then compare the 66% performance thresholds of the model's success rate $p_m(\hat{o}_{\text{correct}}|y_s, y_{h,o}, y_{v,o})$ against the human success rate $p_e(\hat{o}_{\text{correct}}|y_s, y_{h,o}, y_{v,o})$ as reported by Hillis et al. (2002). To set the various model parameters: The prior variance σ_y^2 is fixed globally to an arbitrary large value so as to be fairly uninformative; the prior mean μ_y is assumed known; the unimodal variances σ_h^2 and σ_v^2 and so forth are determined a priori for each experiment and subject by fitting to the unimodal data as in Hillis et al. (2002); and only π_c is fit to the data for each multisensory experiment, with $\pi_c = 0.935$ and 0.99 for the across and within-modality cases, respectively (see Hospedales & Vijayakumar, 2009, for further details).

Detection Threshold Contours Figures 4.10A and 4.10B illustrate the across- and within-modality results, respectively, for the two sample subjects from Figure 4.8. The experimental data (dots) are shown along with the global performance of the model across the whole input space (grayscale background, with white indicating 100% success) and the 66% performance contour (blue lines). The human experimental measurements broadly define a region of nondetection centered about the standard stimuli and slanted along the cues-discordant line and stretched slightly outside the bounds of the inner unimodal threshold rectangle. The extent of the nondetection region along this line is increased somewhat in the

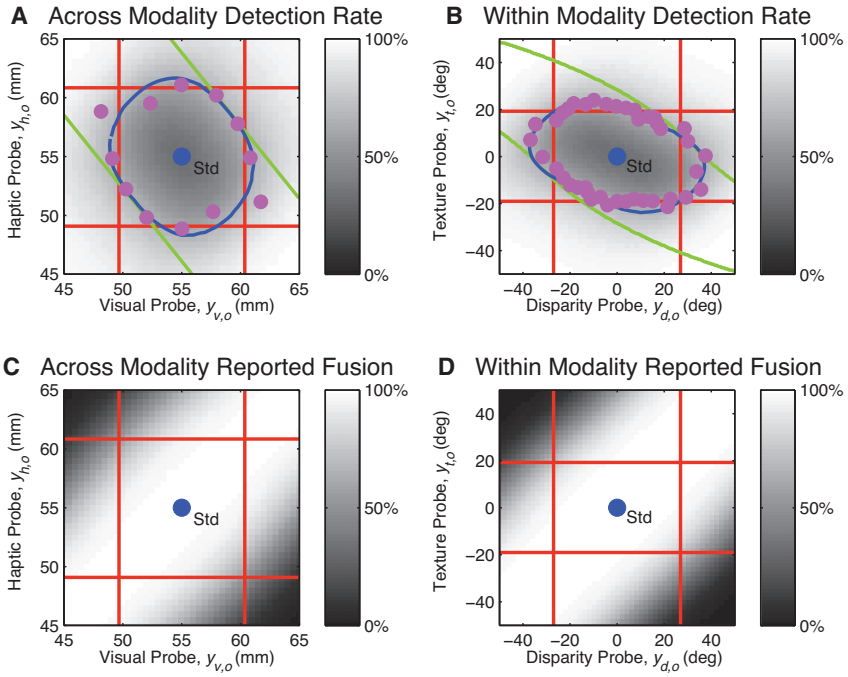

Figure 4.10 Oddity-detection predictions of the structure-inference approach. (A,B) Oddity-detection-rate predictions for the ideal Bayesian observer (grey-scale background) using a variable-structure model (Fig. 4.9); Oddity-detection contours of our model (blue lines) and human (magenta points) are overlaid with the MLI prediction (green lines); Chance = 33%. (C,D) Fusion report rates for the ideal observer using the variable-structure model. Across-modality conditions are reported in (A,C) and within-modality conditions are reported in (B,D).

within-modality case as compared to the across-modality case (Hillis et al., 2002). Recall that the only free parameter varying between these experiments is the common-causation prior π_c, (a larger π_c leads to a longer band of nondetection), which would be expected to vary between pairs of cue modalities.

The MLI model makes the qualitative error of predicting infinite bands of indiscriminability (Fig. 4.10, green lines). In contrast, our Bayesian model provides an accurate quantitative fit to the data (Fig. 4.10, blue lines).

To gain some intuition into these results, consider the normalized distribution of the data (Eq. 4.18) under each model. For example, for $o = 3$, the probability mass in the standard part p_s lies bunched on a four-dimensional line through the standard (where $x_{h,1} = x_{v,1} = x_{h,2} = x_{v,2}$). The probability mass in the probe part p_o is a mixture between a simple model ($C = 1$) around $x_{h,3} = x_{v,3}$, and a more complex model ($C = 0$), spread more uniformly over the space. Therefore, model $o = 3$ will be likely for multisensory observations involving a set of similar pairs $i = 1$, $i = 2$ and a third pair $i = 3$, which is either different from the first set or different from each other.

Perception of Fusion Another benefit of the full generative modeling of this problem is that it also yields a perceptual inference for the fusion (common multisensory source) of the probe $p(C|\{x_{h,i}, x_{v,i}\}_{i=1}^{3})$. This is shown in Figures 4.10C and 4.10D and corresponds to the predicted human answer to the question "*Do you think the odd visual and haptic observations are caused by the same object, or have they become discordant?*" This question was not asked

systematically (Hillis et al., 2002), but they did note that subjects sometimes reported oddity detection by way of noticing the discordance of cues, which is in line with the strategy that falls out of inference with our model.

Along the cues-concordant line, the model has sensibly inferred fusion (Fig. 4.10C and D, quadrants 1 and 3). In these regions, the model can effectively detect the probe (Fig. 4.10A and B, quadrants 1 and 3), and the fused probe estimate $\hat{y}_o$ is different from the standard probe estimate $\hat{y}_s$. Considering instead trials moving away from the standard along the cues-discordant line, the model eventually infers fission (Fig. 4.10C and D, quadrants 2 and 4). The model infers the probe stimuli correctly in these regions (Fig 4.10A and B, quadrants 2 and 4) where the mandatory fusion models cannot (Fig. 4.10a, b, quadrants 2 and 4, green lines) because the probe and standard estimates would be the same $\hat{y}_o = \hat{y}_s$. The strength of discrepancy between the cues required before the fission is inferred depends on the variance of the observations (σ_h^2 and σ_v^2) and the strength of the fusion prior π_c, which will vary depending on the particular subject, combination of modalities, and task.

Discussion

We have developed a Bayesian ideal-observer model for multisensory oddity detection and tested it by reexamining the experiments of Hillis et al. (2002). In those experiments, the standard maximum-likelihood-integration ideal-observer approach failed with drastic qualitative discrepancy compared to human performance; however, we argue that this was due to simple MLI being an inappropriate model rather than a failure of ideal-observer modeling or human suboptimality. The more complete Bayesian ideal-observer model developed here represents the full generative model of the experimental task. This required modeling the multisensory oddity-detection problem as a full model-selection problem with potentially variable probe structure. The Bayesian ideal-observer provides an accurate quantitative explanation of the data with only one free parameter, π_c, which represents a clearly interpretable quantity: prior probability of common causation. Moreover, our interpretation of the problem is satisfying in that it models explicitly and generatively the unknown discrete index o of the odd object: a quantity that the brain is clearly computing since it is the goal of the task.

Generative Modeling Assumptions We have consciously made a stronger assumption than MLI does about how much the human subject knows about the experiment, notably that probe stimulus was possibly discordant. The justification for this is that the subjects were instructed to detect oddity by any means, for which both interstimuli and within-stimulus intercue discrepancy are reasonable indicators. We therefore expect that perceptual circuitry dealing with oddity detection should allow for both kinds of oddity, and as such we model both. Moreover, as discussed in the Introduction, from a normative point of view on generative modeling and ideal observers, we should start with the assumption that the subject has—or learns over the session—a good generative model of the problem; and we were able to model the data without altering this assumption. Of course, it makes more sense for the perceptual system to allow for intermodal discrepancy (because we regularly see and touch different things simultaneously) than intramodal discrepancy as in the texture-disparity case. Nevertheless, this second unintuitive assumption allowed us to make a much better model of the experiment. Exactly why intramodal discrepancy should be permitted and how it is resolved by the perceptual system are open research questions, but we speculate that this could imply some sharing of perceptual integration circuitry between different cue pairs.

Alternative Oddity Models A simple estimator for unimodal three-alternative oddity task is the "triangle rule" (Macmillan & Creelman, 2005). This measures the distances between all three-point combinations, discards the two points with minimum distance between them, and nominates the third point as odd. Note that this simple rule does not provide an acceptable alternative model of the *multisensory* oddity detection scenario studied here because it still does not address the uncertain correspondence

between multisensory observations. Specifically, if the multisensory observations were considered to be fused first (Eq. 4.16), metameric discordant probe observations would still occur—and these cannot be detected by this rule, again producing an infinite band of nondetectability (Fig. 4.8, green lines). In contrast, if the rule were applied directly to the multisensory observations in two dimensions, there would be no room for fusion effects, and detection would be good throughout, in contrast to the tendency toward fusion illustrated by the human data (Fig. 4.8, magenta dots).

Generative Modeling and Structure Inference The theory and practice of generative modeling for inference problems is extensively studied in other related fields, for example, artificial intelligence (Bishop, 2006). In this context, generative modeling of uncertain causal structure in inference tasks goes back to Bayesian multinets (Geiger & Heckerman, 1996). Today, this theory is applied, for example, in building artificial-intelligence systems to explicitly understand "who said what" in multiparty conversations (Hospedales & Vijayakumar, 2008).

Robust Cue Combination A variety of recent studies have investigated the limits of multisensory cue combination and have reported "robust" combination, that is, fusion when the cues are similar and fission when the cues are dissimilar (Bresciani, Dammeier, & Ernst, 2006; Ernst, 2005; Körding, Beierholm et al., 2007; Roach, Heron, & McGraw, 2006; Shams et al., 2005; Wallace et al., 2004). Some authors have tried to understand robust combination by simply defining a correlated joint prior $p(y_h, y_v)$ over the multisensory sources like y_h and y_v (Bresciani et al., 2006; Ernst, 2005, 2007; Roach et al., 2006). These are in general special cases of the full generative approach introduced here (and the equivalent models for other experimental paradigms, e.g., Körding, Beierholm et al., 2007). In the correlated-prior approach, the uncertain structure C, is not represented, and the joint prior over latents is defined as $\sum_C p(y_{h,o}, y_{v,o}|C, \theta)p(C|\theta)$. See Chapter 2 for more details. In our case this would be unsatisfactory because the perceptual system would then not represent causal structure, which subjects do infer explicitly in the work of Hillis et al. (2002) and other related experiments (Wallace et al., 2004). Another reason for the perceptual system to represent and infer causal structure explicitly is that it may be of intrinsic interest. For example, in an audio-visual context, explicit knowledge of structure corresponds to knowledge of "who said what" in a conversation (for example, see Hospedales & Vijayakumar, 2008).

CONCLUSIONS

In this chapter, we have argued that the normative modeling approach of choice for perceptual research should be generative modeling of the perceptual task for each experiment. In this chapter, we have illustrated two sets of experiments in which striking results in human perception can only be explained by full generative models of the respective tasks. These were in domains as diverse as multisensory integration for oddity detection (Hospedales & Vijayakumar, 2009), and visual-proprioceptive integration for sensorimotor adaptation (Haith et al., 2008). The nature of the generative models is quite different in each of these cases: For multisensory integration we considered models in which the unknown variables to be estimated are discrete variables describing the dependency between observations. For sensorimotor learning, we considered a model with continuous, time-varying unknown variables that describe the various possible sources of systematic error affecting each sensory observation. The success of these two contrasting models supports the quite general principle—that the experimental results can only be properly explained by considering a complete generative model of the subject's observations.

In our view, there are two key areas for future research: perceptual learning and physiological implementation. Chapter 9 of this volume introduces some current research progress in perceptual learning. This encompasses questions such as: How do people learn appropriate generative models and parameters for particular tasks? Are there limits to the types of learnable

distributions (e.g., Gaussian, unimodal) and the complexity of learnable models? In online learning, how can the brain adapt parameters online rapidly from trial to trial? How does the brain know when to adapt an existing model or set of parameters versus creating a new one for a new task? Chapter 21 of this volume introduces some current research progress in physiological implementation. This encompasses questions such as: How could these models be computed by biological machinery? Does the brain carry out the exact ideal-observer computations like those we describe here, or is it using heuristics that offer a good approximation in the circumstances considered here. Insofar as human performance falls short of ideal-observer performance in particular experiments, what can this tell us about the architecture of the brain?

ACKNOWLEDGMENTS

T. M. H. and A. M. H. were supported by the UK EPSRC/MRC Neuroinformatics Doctoral Training Center (Neuroinformatics DTC) at the University of Edinburgh. S. V. is supported through a fellowship of the Royal Academy of Engineering in Learning Robotics, cosponsored by Microsoft Research and partly funded through the EU FP6 SENSOPAC and FP7 STIFF projects. We thank Carl Jackson and Chris Miall for assistance with the sensorimotor adaptation experiments.

REFERENCES

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, *20*, 1391–1397.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York, NY: Springer.

Bresciani, J-P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, *6*, 554–564.

Donchin, O., Francis, J. T., & Shadmehr, R. (2003). Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: Theory and experiments in human motor control. *Journal of Neuroscience*, *23*, 9032–9045.

Ernst, M. O. (2005). A Bayesian view on multimodal cue integration. In G. Knoblich, M. Grosjean, I. Thornton, & M. Shiffrar (Eds.), *Human body perception from the inside out* (pp. 105–131). Oxford, England: Oxford University Press.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7(5):7*, 1–14.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Geiger, D., & Heckerman, D. (1996). Knowledge representation and inference in similarity networks and bayesian multinets. *Artificial Intelligence*, *82*, 45–74.

Ghahramani, Z., Wolpert, D. M. & Jordan, M. I. (1997). Computational models for sensorimotor integration. In P. G. Morasso & V. Sanguineti (Eds.), *Self-organization, computational maps and motor control* (pp. 117–147). Amsterdam, The Netherlands: North-Holland.

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*, *15*, 20–29.

Haith, A., Jackson, C., Miall, C., & Vijayakumar, S. (2009). Unifying the sensory and motor components of sensorimotor adaptation. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 21*, 593–600.

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, *298*, 1627–1630.

Hospedales, T., & Vijayakumar, S. (2008). Structure inference for Bayesian multisensory scene understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*, 2140–2157.

Hospedales, T., & Vijayakumar, S. (2009). Multisensory oddity detection as Bayesian inference. *PLoS ONE*, *4*, e4205.

Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict

studies with stereoscopic and figure cues to slant. *Journal of Vision*, *7(7):5*, 1–24.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*, e943.

Körding, K. P., Tenenbaum, J. B., & Shadmehr, R. (2007). The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nature Neuroscience*, *10*, 779–786.

Krakauer, J. W., Mazzoni, P., Ghazizadeh, A., Ravindran, R., & Shadmehr, R. (2006). Generalization of motor learning depends on the history of prior action. *PLoS Biology*, 4e316.

MacKay, D. (2003). *Information theory*, *inference, and learning algorithms.* Cambridge, England: Cambridge University Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Redding, G. M., & Wallace, B. (1996). Adaptive spatial alignment and strategic perceptual-motor control. *Journal of Experimental Psychology*, *22*, 379–394.

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: A strategy for balancing the costs and benefits of audio-visual integration. *Proceedings Biological Sciences*, *273*, 2159–2168.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, *408*, 788.

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, *16*, 1923–1927.

Simani, M. C., McGuire, L. M., & Sabes, P. N. (2007). Visual-shift adaptation is composed of separable sensory and task-dependent effects. *Journal of Neurophysiology*, *98*, 2827–2841.

van Beers, R. J., Wolpert, D. M., & Haggard, P. (2002). When feeling is more important than seeing in sensorimotor adaptation. *Current Biology*, *12*, 834–837.

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, *158*, 252–258.