

# CHAPTER 12

## Multisensory Perception: From Integration to Remapping

*Marc O. Ernst and Massimiliano Di Luca*

---

### INTRODUCTION

The brain receives information about the environment from all the sensory modalities, including vision, touch, and audition. To interact efficiently with the environment, this information must eventually converge to form a reliable and accurate multimodal percept. This process is often complicated by the existence of noise at every level of signal processing, which makes the sensory information derived from the world unreliable and inaccurate. We define *reliability* as the inverse variance of the probability distribution that describes the information a sensory signal contributes to the perceptual estimation process. In contrast, *accuracy* is defined as the probability with which the sensory signal truly represents the magnitude of the real-world physical property that it reflects. In other words, it is inversely related to the probability of a sensory signal being biased with respect to the world property. There are several ways in which the nervous system may minimize the negative consequences of noise in terms of reliability and accuracy. Two key strategies are to combine redundant sensory estimates and to use prior knowledge. There is behavioral evidence that the human nervous system employs both of these strategies to reduce the adverse effects of noise and thus to improve perceptual estimates.

In this chapter, we elaborate further on how these strategies may be used by the nervous

system to obtain the best possible estimates from noisy signals. We first describe how weighted averaging can increase the reliability of sensory estimates, which is the benefit of multisensory integration. Then, we point out that integration can also come at a cost of introducing inaccuracy in the sensory estimates. This shows that there is a need to balance the benefits and costs of integration. This is done using the Bayesian approach, with a joint likelihood function representing the reliability of the sensory estimates (e.g.,  $\hat{S}_V$  and  $\hat{S}_H$ , for visual and haptic sensory estimates) and a joint prior probability distribution providing the co-occurrence statistics of sensory signals  $p(S_V, S_H)$ , that is, the prior probability of jointly encountering an ensemble of sensory signals derived from the world. This framework naturally leads to a continuum of integration between fusion and segregation. We further show how this framework can be used to model the breakdown of integration by having the joint prior conditioned on multisensory discordance (i.e., a separation of the sensory signals in time, space, or some other measure of similarity). If the multisensory signals differ constantly over a period of time, because they may be consistently inaccurate, recalibration of the multisensory estimates will be the result. The rate of recalibration can be described using a Kalman-filter model, which can also be derived from the Bayesian approach. We conclude by

proposing how integration and recalibration can be jointly described under this common approach.

## MULTISENSORY INTEGRATION

For estimating a specific environmental property, such as the size of an object in the world  $S_W$ , there are often multiple sources of sensory information available. For example, an object's size can be estimated by sight and touch (haptics),  $S_V$  and  $S_H$ . Typical models of sensory integration assume unbiased (accurate) sensory signals (i.e.,  $S_V = S_H$ ) with normally distributed noise sources that are independent, a situation in which sensory integration is beneficial (see Chapter 1; Landy, Maloney, Johnston, & Young, 1995). For the estimation of an object's size from vision and touch, the assumption of independent noise sources is likely to be true since most of the neuronal processing for sensory signals, that is, their transmission from sensory transducers to the brain, is largely independent. As was introduced in Chapter 1, Figure 12.1 illustrates the optimal mechanism of sensory combination given these assumptions and given that the goal is to compute a minimum-variance estimate. This can be considered the standard model of sensory integration. The likelihood functions represent two independent estimates of size, the visual

size estimate  $\hat{S}_V$  and the haptic size estimate  $\hat{S}_H$ , based on sensory measurements ( $z_V, z_H$ ) that are corrupted by noise (with standard deviations  $\sigma_V$  and  $\sigma_H$ ). The integrated multisensory estimate  $\hat{S}_{VH}$  is a weighted average of the individual sensory estimates with weights  $w_V$  and  $w_H$  that sum up to unity (Cochran, 1937):

$$\hat{S}_{VH} = w_V \hat{S}_V + w_H \hat{S}_H, \text{ where } w_V + w_H = 1. \quad (12.1)$$

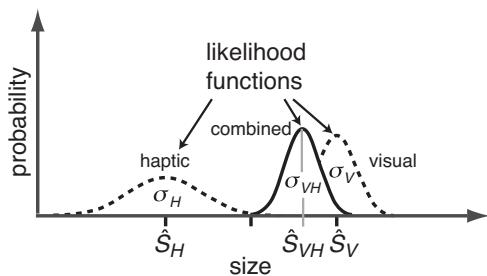
To achieve optimal performance, the chosen weights need to be proportional to the reliability  $r$ , which is defined as the inverse of the signal variance:

$$w_j = \frac{r_j}{\sum_i r_i}, \text{ with } r_i = \frac{1}{\sigma_i^2}. \quad (12.2)$$

The indices  $i$  and  $j$  refer to the sensory modalities ( $V, H$ ). The modality that provides more reliable information in a given situation is given a higher weight, and so has a greater influence on the final percept. In the example shown in Figure 12.1, visual information about the size of the object is four times more reliable than the haptic information. Therefore, the combined estimate (the weighted sum) is "closer" to the visual estimate than the haptic one (in the present example the visual weight is 0.8 according to Eq. 12.2). In another circumstance where the haptic modality might provide a more reliable estimate, the situation would be reversed.

Given this weighting scheme, the benefit of integration is that the variance of the combined estimate from vision and touch is less than that of either of the individual estimates that are fed into the averaging process. Therefore, the combined estimate arising from integration of multiple sources of independent information shows greater reliability and diminished effects of noise. Mathematically, this is expressed by the combined reliability  $r$  being the sum of the individual reliabilities:

$$r = \sum_i r_i. \quad (12.3)$$



**Figure 12.1** Schematic representation of the likelihood functions of the individual visual and haptic size estimates  $\hat{S}_V$  and  $\hat{S}_H$  and of the combined visual-haptic size estimate  $\hat{S}_{VH}$ , which is a weighted average according to Eq. 12.1. The variance associated with the visual-haptic distribution is less than either of the two individual estimates (Eq. 12.3). (Adapted from Ernst & Banks, 2002.)

Given that all estimates are unbiased, this integration scheme can be considered statistically optimal, since it provides the lowest possible variance of its combined estimate. Thus, this form of sensory combination is the best way to reduce uncertainty given the assumptions that all estimates are accurate and contain Gaussian-distributed, independent noise (Chapter 1). Even if the noise distributions of the individual signals displayed a correlation, averaging of sensory information would still be advantageous and the combined estimate would still be more reliable than each individual estimate alone (Oruç, Maloney, & Landy, 2003).

Several studies have tested this integration scheme empirically (e.g., Gharahmani, Wolpert, & Jordan, 1997; van Beers, Sittig, & Denier van der Gon, 1998, 1999). In 2002, Ernst and Banks showed that humans integrate visual and haptic information in such a statistically optimal fashion. It has further been demonstrated that this finding of optimality also holds across and within other sensory modalities, for example, vision and audition (e.g., Alais & Burr, 2004; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003; Landy & Kojima, 2001). Thus, weighted averaging of sensory information appears to be a general strategy employed by the perceptual system to decrease the detrimental effects of noise.

If redundant sources of sensory information are absent or if the noises of these sources are perfectly correlated, averaging different estimates is not an option to reduce noise. However, because the world is structured quite regularly, the nervous system can use prior knowledge about such statistical regularities to reduce the uncertainty and ambiguity in neuronal signals. Prior knowledge can also be formalized as a probability distribution in a manner similar to that for sensory signals corrupted by noise. For example, let us consider the distribution of velocities for all objects. While some objects in our environment do move around occasionally, from a purely statistical point of view, on average most objects are likely to remain stationary at most times, that is, the velocity of an object is most likely to be zero. Thus, a reasonable probability distribution

describing the velocity of all objects is centered at zero with some variance (Stocker & Simoncelli, 2006; Weiss, Simoncelli, & Adelson 2002). This prior knowledge can be combined with unreliable sensory evidence in order to minimize the uncertainty in the final velocity estimate. If all the probability distributions are Gaussian, using Bayes' rule it is possible to derive that the combined posterior estimate (the maximum a posteriori or MAP estimate) is a weighted average as well; however, now it is a weighted average between the prior and the likelihood function, that is, the sensory evidence:

$$\hat{S}_{MAP} = w_{likelihood} \hat{S}_{likelihood} + w_{prior} \hat{S}_{prior} \quad (12.4)$$

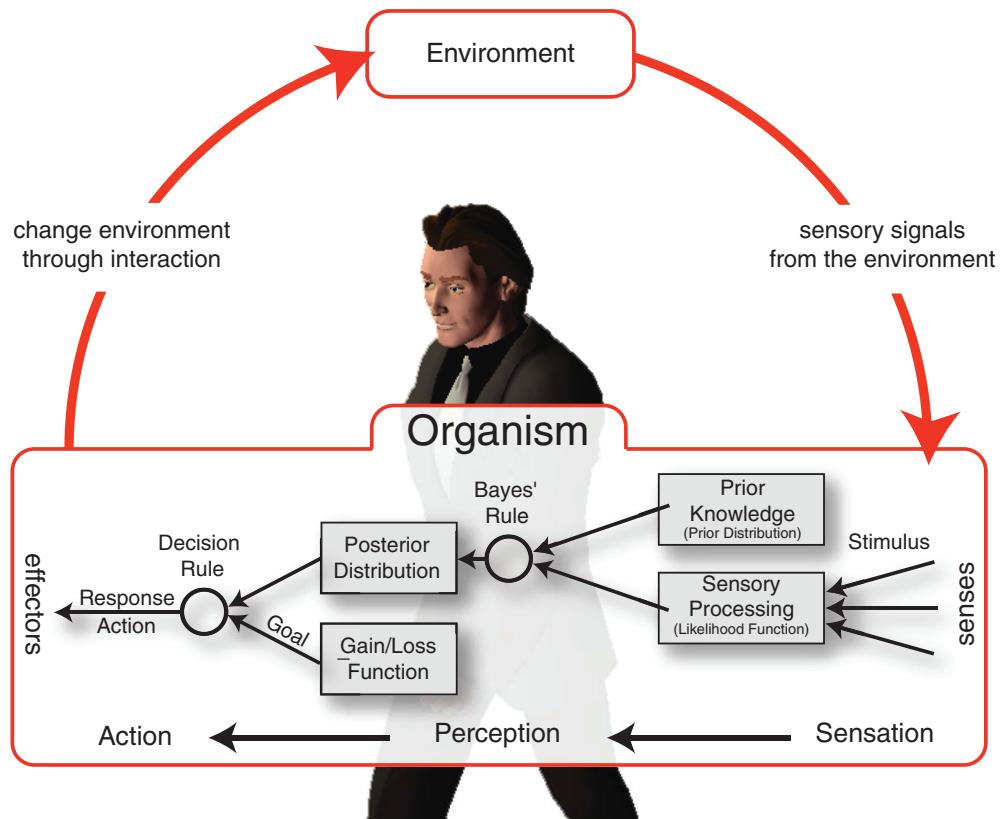
The reliability of the MAP estimate then is given by:

$$r_{MAP} = r_{likelihood} + r_{prior}. \quad (12.5)$$

The principles of weighted averaging and the use of prior knowledge can be combined and placed into a larger mathematical framework of optimal statistical estimation and decision theory, known as Bayesian decision theory (Chapter 1; Mamassian, Landy, & Maloney, 2002). This approach is illustrated in Figure 12.2 in the context of the action-perception loop. Psychophysical experiments have confirmed that at least some aspects of human perception and action that deal with noise and uncertainty can be described well using this Bayesian framework (e.g., Adams, Graf, & Ernst, 2004; Kersten, Mamassian, & Yuille, 2004; Körding & Wolpert, 2004; Stocker & Simoncelli, 2006).

## THE COST OF INTEGRATION

While weighted averaging of sensory measurements or use of prior knowledge has the benefit of reducing noise and uncertainty in perceptual estimates, it also incurs a potential cost. The cost is the introduction of potential biases into perception. Biases can occur, for example, when the sensory estimates ( $\hat{S}_V$  and  $\hat{S}_H$ ) as defined by the likelihood functions and thus sensory signals ( $S_V$  and  $S_H$ ) do not accurately represent the physical stimuli ( $S_{W_V}$  and  $S_{W_H}$ ). Accuracy of



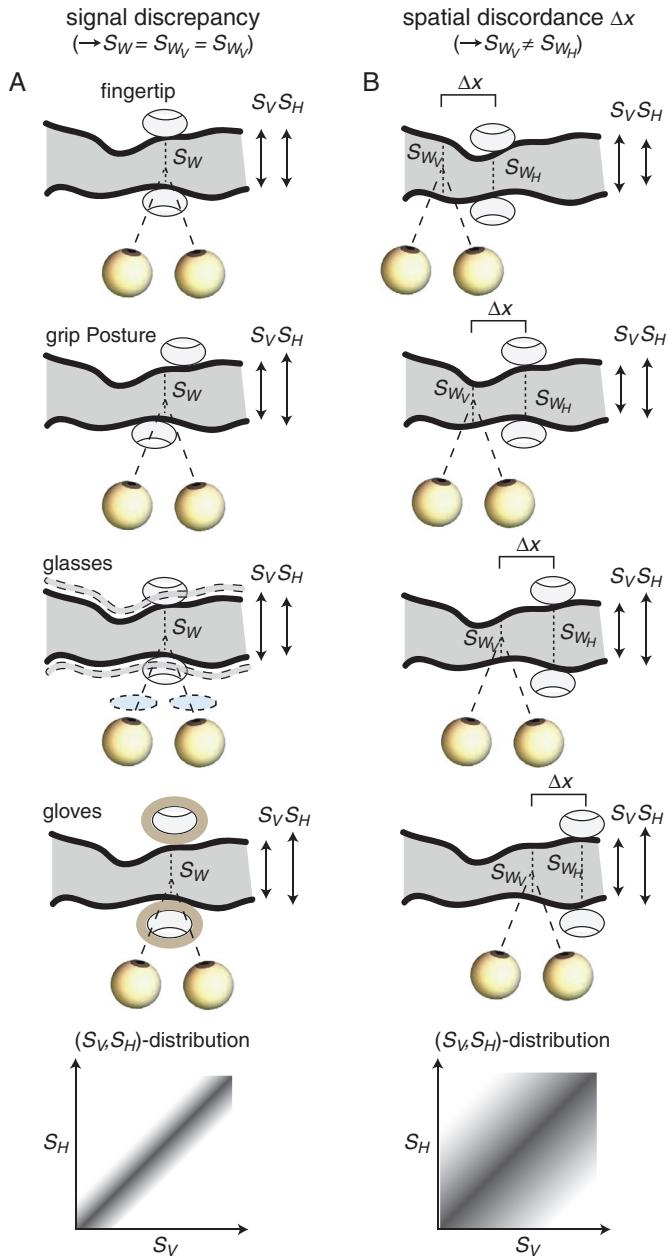
**Figure 12.2** The action/perception-loop schematically illustrates the processing of information according to Bayesian decision theory. Multiple sensory signals are averaged during sensory processing and then combined with prior knowledge, to derive the most reliable, unbiased estimate (posterior) that can be used in a task that has a goal as defined by a gain or loss function. (Adapted from Ernst & Bülthoff, 2004.)

the sensory estimates was one of the assumptions made in the previous section for deriving the optimal integration scheme (Chapter 1). However, if the estimates are no longer accurate due to external or internal influences on the signals, the potential cost of biases has to be considered.<sup>1</sup> Examples of sources of inaccuracies in signals may be muscle fatigue, variance in grip posture, or wearing gloves. Additionally there might be glasses that distort the visual image and so affect visual position estimates, or effects of temperature or humidity that affect sound

<sup>1</sup>Throughout the paper we are only considering additive biases, although the general scheme can be extended to other forms of biases, for example, multiplicative biases.

propagation and thus affect auditory estimates, to name just a few. Figure 12.3A illustrates some examples of processes that might affect the accuracy of visual-haptic size estimates. The top panel shows sensory signals ( $S_V$  and  $S_H$ ) that are accurate with respect to the world property  $S_W$  (so  $S_W = S_{W_V} = S_{W_H} = S_V = S_H$ ) to be estimated (i.e., the size of the object at a specific position), followed by three examples of  $S_V$  and  $S_H$  signals that are inaccurate and contain an additional bias  $B$  (i.e.,  $S_V = S_{W_V} + B_V$  and  $S_H = S_{W_H} + B_H$ ). For now we assume that the signals are derived from the same location, so the visual and haptic sizes to be estimated are identical:  $S_W = S_{W_V} = S_{W_H}$ .

If the sensory signals  $S_V$  and  $S_H$ , and hence the sensory estimates derived from these signals,



**Figure 12.3** (A) Visual and haptic size signals  $S_V$  and  $S_H$  measured near the same location on an object at which the true size is  $S_W$ . In this case visual and haptic sizes are identical ( $S_{W_V} = S_{W_H}$ ). The sensory signals can be corrupted by various disturbances, which affect their accuracy, such as different grip postures, glasses, or gloves. (B) Visual and haptic size signals  $S_V$  and  $S_H$  derived from locations on an object in close proximity (offset horizontally by  $\Delta x$ ). In this case visual and haptic sizes may differ slightly ( $S_{W_V} \neq S_{W_H}$ ). Thus, the visual and haptic size signals will also differ slightly due to variations in the shape of the object. However, in general there will still be a correlation between the  $S_V$  and  $S_H$  signals as the object's size varies smoothly. Most probably this correlation will decrease with increasing  $\Delta x$ . In both cases, the lower panel labeled  $(S_V, S_H)$ -distribution provides the co-occurrence statistics of the signal values  $S_V$  and  $S_H$  that build the basis for the prior used for multisensory integration.

$\hat{S}_V$  and  $\hat{S}_H$ <sup>2</sup> are inaccurate, that is, if they are biased by  $B = (B_V, B_H)$  with respect to the world property  $S_W$  or with respect to each other (sensory discrepancy  $D = S_V - S_H = (S_W + B_V) - (S_W + B_H)$ ), their respective values need not necessarily agree even when they are derived from the same location ( $S_W = S_{W_V} = S_{W_H}$ ). In such a case, weighted averaging of the estimates derived from these biased signals will inevitably also bias the combined estimate. To avoid the cost of biased estimates, the perceptual system must be able to infer how accurate the signals are. *This is a difficult problem that cannot be determined directly from the sensory estimates, because these estimates do not carry information about their own accuracy.* Reliability, on the other hand, which is the inverse variance associated with the estimates, can be directly assessed from sensory measurements. Furthermore, the mere existence of a discrepancy between sensory estimates  $\hat{D} = \hat{S}_V - \hat{S}_H$  does not reveal whether some of the estimates are inaccurate, because

<sup>2</sup>Variables with a hat always denote noisy sensory estimates, whereas variables without a hat represent world signals from which the sensory estimates are derived.

even when they are accurate, the presence of noise in the estimation process will cause the respective peaks of their likelihood functions to disagree slightly (as illustrated in Fig. 12.1). We will discuss later in the chapter how persistently biased estimates may be avoided through the process of recalibration.

A problem regarding potential biases also exists while using prior knowledge to reduce perceptual uncertainty. If the prior probability distribution does not *accurately* describe the statistics of the current environment and if the mean of the prior distribution differs from the mean of the sensory measurements, it will introduce a bias in the final perceptual estimates. Evidence for this phenomenon can be found in several perceptual illusions, for instance, the one illustrated in Figure 12.4. Both pictures show footprints in the sand. However, most people see the left image as an indentation in the sand, whereas they see the right image as if it were embossed or raised from the surface. The reason for this counterintuitive perception is the inherently ambiguous nature of the image and the need to make certain prior assumptions in order to interpret it (Rock, 1983). The prior assumption we make in this case is that the light



**Figure 12.4** Effect of the light-from-above prior on perception using ambiguous images. The left and right images show footprints in the sand. In the left image the light illuminating the scene is actually coming from above, and the footprint is correctly seen as an indentation. In the right image, which is the left image presented upside down, the light is coming from below. Employing the light-from-above prior in this situation causes the footprint to be seen as embossed or raised from the surface.

source in the image is placed above the surface (Brewster, 1826; Mamassian & Goutcher, 2001). The assumption reflects our common world experience of always having artificial or natural light from above (Dror, Willsky, & Adelson, 2004). In the illustration, this assumption is only correct for the left image. The illusion arises when one views the right image. In the right image the footprints are actually illuminated from below. Thus, making prior assumptions about light from above, that is, using an inappropriate prior for the current situation, forces our perception toward a bias that causes us to see the footprints raised from the surface.

To interact successfully with the environment in order to, say, point to an object, *the goal of the sensorimotor system must be to derive accurate estimates for the motor actions to be performed*. For example, we might wish to interact with the environment by touching one of the toes of the footprints shown in Figure 12.4. Evoking an inappropriate prior will introduce a bias into the inferred depth used for pointing. That is, in the right part of Figure 12.4 we would wrongly point to the illusory perceived embossed toe instead of the actual imprinted toe (Hartung, Schrater, Bülthoff, Kersten, & Franz, 2005). Therefore, biases such as those discussed earlier are undesirable and should be avoided. This, in turn, predicts that multisensory integration must break down with an increase of conflicting information between the multisensory sources. For this reason prior knowledge should be disregarded if it is evident that the sensory information is derived from an environment with statistical regularities that conflict with those represented by the prior probability distributions. There is experimental evidence to back up both these claims, which will be discussed next.

As indicated earlier, there are many perceptual illusions that arise because prior assumptions bias the percept. Another example is the use of prior knowledge about symmetry or isotropy in visual slant perception (Palmer, 1985). When asked for the three-dimensional interpretation of an ellipse, humans consistently see the ellipse as a circle slanted in depth. This perceptual effect

is explained using a prior for symmetry which when evoked interprets the ellipse as a circle. This may make sense because, considering the statistics of our world, we are more likely to encounter circles than ellipses. Therefore, under these statistical considerations, for the unlikely event that the ellipse is really an ellipse, this prior will give rise to a biased percept. Knill (2007a) showed that we down-weight such prior knowledge for seeing circles if we are placed in an environment where ellipses or irregular shapes occur more frequently. This is consistent with the idea that we begin to ignore the prior when there is statistical evidence against the symmetry assumption. As a consequence, this strategy saves us from acquiring biases based on false prior assumptions. Along the same lines, Adams et al. (2004) showed that the light-from-above prior (as demonstrated in Fig. 12.4) adapts when observers are put in an environment where the light source is placed predominantly to the left or right, instead of above.

There is also empirical evidence for biases in multisensory perception and for the breakdown of multisensory integration with large discrepancies between the sensory estimates. For example, multisensory integration has been studied experimentally by the deliberate introduction of small discrepancies between sensory signals such that the perceptual consequences of integration are evident in a bias resulting from weighted averaging; a method termed “perturbation analysis” (Young, Landy, & Maloney, 1993). Some notable demonstrations of multisensory biases induced by weighted averaging include shifts in perceived location (Alais & Burr, 2004; Bertelson & Radeau, 1981; Pick, Warren, & Hay, 1969; Welch & Warren, 1980), perceived rate of a rhythmic stimulation (Bresciani, Dammeier, & Ernst 2006, 2008; Bresciani & Ernst, 2007; Bresciani et al., 2005; Gebhard & Mowbray, 1959; Myers, Cotton, & Hilp, 1981; Recanzone, 2003; Shams, Kamitami, & Shimojo, 2002; Shipley, 1964; Welch, DuttonHurt, & Warren, 1986), or perceived size (Ernst & Banks, 2002; Helbig & Ernst, 2007). With larger experimentally induced discrepancies between the perceptual estimates, however, the integration and weighted averaging process breaks down (Knill, 2007b). Integration

breaks down even more rapidly if there is additional evidence that the sources of information do not originate from the same object or event. For example, Gepshtein, Burge, Banks, and Ernst (2005) showed that visual and haptic size integration breaks down rapidly if the visual and haptic information do not come from the same location. That is, location information is used in addition to determine whether to integrate the size estimates. Several studies have shown this breakdown of integration with spatial discordance in a similar way (Jack & Thurlow, 1973; Jackson, 1953; Warren & Cleaves, 1971; Witkin, Wapner, Leventhal, 1952; but see also Recanzone, 2003). The breakdown also happens with temporal discrepancies (e.g., Bresciani et al., 2005; Radeau & Bertelson, 1987; Shams et al., 2002; van Wassenhove, Grant, & Poeppel, 2007). This breakdown of integration with increasing discordance in space and time defines the spatial and temporal windows of integration. It is more generally referred to as robustness of integration.

We have now identified two competing goals of the perceptual-motor system: the first goal, discussed in the previous section, was to achieve the most reliable estimates possible; the second goal, discussed in this section, was to avoid inaccuracy of the estimates, that is, to achieve the most accurate estimates. *To maximize the gain from integration, these two competing goals must be best balanced.* For this the precision (reliability) and accuracy of the sensory estimates has to be known to the system. As mentioned earlier, reliability can in principle be determined online from analyzing the estimates. However, there is no direct information in the sensory signals or estimates that would allow one to determine their accuracy. In the following we will therefore concentrate on the question of how the brain determines whether sensory signals and estimates are accurate, whether there is a discrepancy between the sensory estimates, and so whether to integrate. The same question arises for the use of prior knowledge as well and whether it conforms to the statistics of the present environment. To keep matters simple, however, from now on we will concentrate on the first question.

## BALANCING BENEFITS AND COSTS

Whether to integrate different multisensory estimates depends on the presence of an actual difference  $D$  between the multiple sensory signals. The perceptual system, however, does not have direct access to the sensory signals but only to the estimates derived from these signals. Thus, to estimate what constitutes an actual difference  $D$  between the signals is a question that is itself shrouded in uncertainty because of the noise in the estimation process. That is, when we make estimates  $\hat{D}$  of sensory discrepancies, we are unable to do so reliably because of the noise in such estimates (see Wallace et al., 2004). For this reason, it is practically impossible to determine an absolute threshold for whether to integrate. Every time a discrepancy is detected between two estimates, the perceptual system must determine (either implicitly or explicitly) the reason for such a discrepancy. If the discrepancy  $\hat{D}$  arises from random noise in the processing of the neuronal signals, the discrepancy changes randomly from trial to trial. In this case, by integrating the two estimates, the perceptual system could average out the influence of such noise as shown in the beginning of this chapter. However, if the discrepancy in the estimates  $\hat{D}$  were due to a systematic difference  $D$  between the signals, then the best strategy would require the perceptual system to not integrate the multisensory information. This may occur, for example, in a scenario where the sensory signals to be combined show some inaccuracy (in form of an additive bias  $B$ ) with respect to the world (i.e.,  $S_V = S_{W_V} + B_V$  or  $S_H = S_{W_H} + B_H$ ), or with respect to one another (i.e.,  $D = S_V - S_H$ ). Figure 12.3A illustrates this with a few examples showing how the sensory signals ( $S_V$  and  $S_H$ ) may become inaccurate with respect to the world property  $S_W = (S_{W_V}, S_{W_H})$  to be estimated. As a consequence, determining the reason for the discrepancies in the sensory estimates is a credit-assignment problem with two possibilities: The reason for the discrepancy could either be a difference between the signals or a random perturbation as a result of noise, where both possibilities are uncertain. Since both

possibilities are plausible and have associated uncertainty, the optimal strategy would be to use them both and weight each according to its relative certainty. We call this optimal because it balances the benefit of multisensory integration while minimizing the potential costs associated with it. This intuitive concept forms the basis of a model that we discuss further in the next section.

### **MODELING FUSION, PARTIAL FUSION, AND SEGREGATION**

To summarize, no matter how small it may be, a discrepancy *always* exists between perceptual estimates derived from different signals ( $\hat{D} = \hat{S}_V - \hat{S}_H$ ). Such discrepancies could either be caused by random noise in the estimates (with standard deviations  $\sigma_V$  and  $\sigma_H$ ), which is unavoidable and always present, or it could be caused by a systematic difference  $D$  in magnitude between the sensory signals. To make the best possible use of such discrepant information, the brain must use different and antithetical strategies for random noise and systematic difference. Information should be fused if the discrepancy was caused by random noise in the estimates, and it should be segregated if the discrepancy was caused by an actual difference in the signals. Interestingly, the very determination of the source of the discrepancy, random or systematic, is itself uncertain and difficult to estimate and so the reason for any discrepancy can only be determined with uncertainty. Thus, the best solution to model such a process is to use a fully probabilistic approach.

While our nervous system is capable of processing many complex signals and sources at once, we try to keep matters simple here by considering a discrepancy between only two estimates, each of which represents a property  $S_W = (S_{W_V}, S_{W_H})$  specified by sensory signals  $S = (S_V, S_H)$ . Thus, it is reasonable to think of the integration process in a 2D space (Fig. 12.5), although the problem can be extended easily to higher dimensions. We now continue with the example we used earlier (Eq. 12.1) in which visual and haptic estimates are combined to determine the size of an object  $S_W = (S_{W_V}, S_{W_H})$

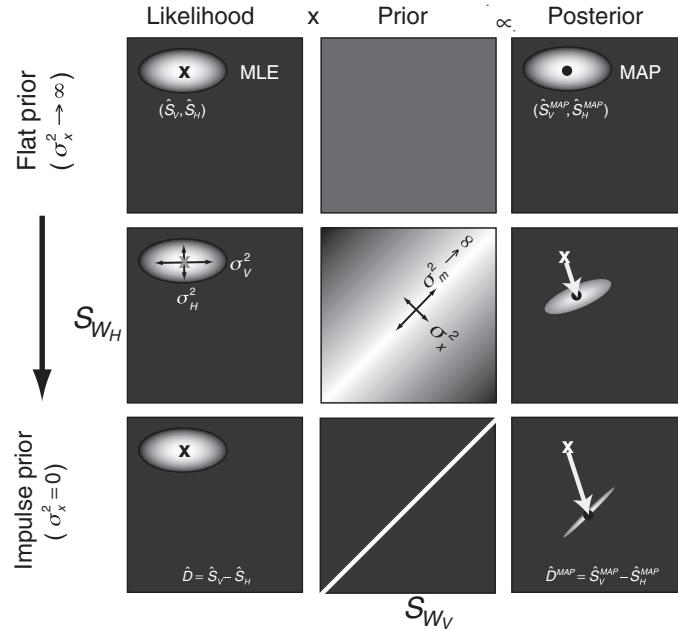
with  $S_{W_V} = S_{W_H}$ . Let  $S = (S_V, S_H)$  be the sensory signals derived from the world  $S_W = (S_{W_V}, S_{W_H})$ , which may be biased ( $B = (B_V, B_H)$ ; Fig. 12.3) with respect to some world property or with respect to one another, and let  $z = (z_V, z_H)$  be the sensory measurements derived from  $S$ . Both the visual and haptic measurements are corrupted by independent Gaussian noise with variance  $\sigma_i^2$ , so  $z_i = S_i + \varepsilon_i$  with  $i$  referring to the individual sensory modality ( $V, H$ )<sup>3</sup>. With these assumptions, the joint likelihood function takes the form of a Gaussian density function:

$$p(z|S) = N(\hat{S}, \Sigma_z) \text{ with } \Sigma_z = \begin{pmatrix} \sigma_V^2 & 0 \\ 0 & \sigma_H^2 \end{pmatrix}, \quad (12.6a)$$

which is a bivariate normal distribution with mean  $\hat{S} = (\hat{S}_V, \hat{S}_H) = (z_V, z_H)$  (i.e., the maximum-likelihood estimates of the sensory signals equal the noisy measurements) and covariance matrix  $\Sigma_z$  (left column in Fig. 12.5).

The likelihood function represents the sensory measurements on a given trial. The goal of this task will be the estimation of a property of the world, such as  $S_W$ , while taking into account both sensory imprecision (due to random noise) and inaccuracy (additive bias). In the rest of this chapter, we will develop a Bayesian model of this process that proceeds in two steps. In the first stage, discussed in this section and the next, we describe how the observer can use Bayes' rule to calculate a posterior distribution of the sensory signals given the noisy measurements,  $P(S_V, S_H|z_V, z_H)$ , and MAP estimates of those sensory signals,  $\hat{S}^{MAP}$ , that take into account prior knowledge of the correlations between the signals,  $p(S_V, S_H)$ . In subsequent sections, we describe how the observer can use prior knowledge of the likely inaccuracy in each modality ( $B_V, B_H$ ) along with current estimates of the discrepancy between

<sup>3</sup>As previously, we assume the visual and haptic estimates are normally distributed and statistically independent. Oruç et al. (2003) and Ernst (2005) describe an analysis of how such a system behaves in case the estimates are not independent and how this may give rise to negative weights.



**Figure 12.5** The combination of visual and haptic measurements with different prior distributions. (Left column) Likelihood functions resulting from noise with standard deviation  $\sigma_V$  twice as large as  $\sigma_H$ ;  $x$  indicates the maximum-likelihood estimate (MLE) of the sensory signals  $\hat{S} = (\hat{S}_V, \hat{S}_H)$ . (Middle column) Prior distributions with variance  $\sigma_m^2 \rightarrow \infty$ , but different variances  $\sigma_x^2$ . Top: flat prior  $\sigma_x^2 \rightarrow \infty$ ; middle: intermediate prior  $0 < \sigma_x^2 < \infty$ ; bottom: impulse prior  $\sigma_x^2 = 0$ . (Right) Posterior distributions, which are the normalized product of the likelihood and prior distributions. A dot indicates the maximum a posteriori (MAP) estimate  $\hat{S}^{MAP} = (\hat{S}_V^{MAP}, \hat{S}_H^{MAP})$ . Arrows correspond to bias in the MAP estimate relative to the MLE estimate. The orientation of the arrows indicates the weighting of the  $\hat{S}_V$  and  $\hat{S}_H$  estimates. The length of the arrow indicates the degree of fusion. (Adapted with permission from Ernst, 2007. Copyright ARVO.)

sensory signals ( $\hat{D}^{MAP} = \hat{S}_V^{MAP} - \hat{S}_H^{MAP}$ ) after integration occurred to solve iteratively the credit-assignment problem: What portion of  $\hat{D}^{MAP}$  should be attributed to the bias  $B_i$  or the world property  $S_{W_i}$  of each modality? The solution of this problem will allow the observer to remap each modality, as a means of providing the best possible (low bias and low uncertainty) estimate of  $S_W$ .

To begin, we assume that the system has acquired a priori knowledge about the probability of jointly encountering a combination of sensory signals encoded in the prior  $p(S_V, S_H)$ . Some examples of visual and haptic signals to size  $(S_V, S_H)$  that might be encountered in conjunction when trying to estimate the world property  $S_W$  are provided in Figure 12.3.

The lower row in Figure 12.3 shows what such a distribution of jointly encountered signals might look like. Figure 12.3 A shows cases where the signals are derived from the same location for which we can assume that  $S_{W_V} = S_{W_H}$ . All these examples show signals with varying accuracy ( $B_i = S_{W_i} - S_i$ ). The point here is that the variance in the joint distribution and hence the variance of the prior learned from these signals is affected by the variability in accuracy of the two signals. Figure 12.3B illustrates a similar example of co-occurrence of visual and haptic signals, but here these signals are derived from slightly disparate locations  $\Delta x$  for which in general  $S_{W_V} \neq S_{W_H}$ . We return to this example in a later section of this chapter when we discuss the link between integration and remapping.

Assuming for now that all the joint distributions are Gaussians, a prior that fulfills what we have discussed thus far can be defined as:

$$p(S) = p(S_V, S_H) = N(n, \Pi)$$

with  $\Pi = R^T \begin{pmatrix} \sigma_m^2 \rightarrow \infty & 0 \\ 0 & \sigma_x^2 \end{pmatrix} R$  (12.7)

which is a bivariate normal distribution with mean  $n = (0, 0)$  and covariance matrix  $\Pi$ .  $\sigma_m^2$  and  $\sigma_x^2$  are the variances of the prior along its principal axes and  $R$  is an orthogonal matrix that rotates the coordinate system by  $45^\circ$  so that the prior is aligned with the diagonal where  $S_V = S_H$  (Fig. 12.5, middle column). We choose the variance along the positive diagonal to be  $\sigma_m^2 \rightarrow \infty$ , which indicates that the probability of jointly encountering two signals ( $S_V, S_H$ ) is independent of their mean value.<sup>4</sup> The second variance,  $\sigma_x^2$ , indicates the spread of the joint distribution, which represents the a priori distribution of possible discrepancies between the signals. Therefore, the probability that the source of any detected discrepancy  $\hat{D}$  is not random noise but an actual difference between the signals  $D = S_V - S_H$  is a function of the variance  $\Pi$  (i.e.,  $\sigma_x^2$ ) of this prior. The diagonal with  $S_V = S_H$  represents the mapping between the signals since it provides the functional relationship between the two. We can therefore also refer to  $\sigma_x^2$  as the *mapping uncertainty*. Furthermore, this distribution also provides a measure of redundancy between the two signals; the smaller the variance  $\sigma_x^2$ , the more redundant the signals are with respect to one another.

Figure 12.5 illustrates three examples of the model described earlier for prior distributions with different  $\sigma_x^2$  (middle column) ranging from very large (top row) to near zero (bottom row). A prior probability with  $\sigma_x^2 \rightarrow \infty$  corresponds to a state in which any possible combination of  $S_V$  and  $S_H$  signals contains roughly an equal a priori probability of occurrence. Such a prior is often referred to as a “flat prior.” In this extreme case

of  $\sigma_x^2 \rightarrow \infty$ , there is no mapping between the sensory signals or estimates derived from them and thus the discrepancy between the estimates is ill defined. Theoretically, however, one might argue that the accuracy of the signals with respect to this ill-defined mapping approaches zero. This has also been referred to as signals that are invalid (with respect to the property defined by the mapping). Such a situation is an example of signals  $S_V$  and  $S_H$  that do not carry redundant information. Thus, as an example we could take any set of nonrelated signals, such as the luminance and the stiffness of an object, which are highly unlikely to carry any redundant information (Ernst, 2007) and can co-occur in any possible combination.

A prior probability with  $\sigma_x^2 = 0$ , on the other hand, corresponds to a state in which signals occur only for the condition  $S_V = S_H$ . Such a prior relates to signals that are always perfectly accurate (with respect to the property of interest). In this situation the prior probability of encountering an actual difference  $D$  between the signals is zero. Thus, in this situation the sensory signals are completely redundant. While such a scenario would be purely theoretical because there is always some variance present, indirect empirical evidence that humans use very tight priors was provided by Hillis, Ernst, Banks, and Landy (2002), who found close to mandatory fusion of disparity and texture estimates to slant (see later discussion).

An intermediate value of  $\sigma_x^2$  corresponds to a state in which the probability distribution indicates some uncertainty with respect to the possible co-occurrence of signal values  $S_V$  and  $S_H$ . Such a prior relates to signals that display some inaccuracy with respect to the mapping and thus there exists a nonzero probability of encountering various differences  $D$  between the signals. The signals in this situation are thus only *partially* redundant (with respect to one another). Since this prior refers to the probability of co-occurrence of certain signals, that is, it represents the prior probability of jointly encountering an ensemble of sensory estimates, in earlier work this prior  $p(S_V, S_H)$  has also been referred to as the “coupling prior” (Bresciani et al., 2006; Ernst, 2005, 2007). Realistically, all

<sup>4</sup>Thus,  $n$  could have any value with  $S_V = S_H$ . We arbitrarily choose  $n = (0, 0)$ .

cases of multisensory integration, such as size estimation from vision and touch, fall into this category (Ernst, 2005; Hillis et al., 2002). This is because there is always some probability that the signals are inaccurate due to external or internal factors, such as muscle fatigue, optical distortion, or other environmental or bodily influences (Fig. 12.3A).

Using Bayes' rule (see Chapter 1), the joint likelihood function obtained from the sensory signals is combined with prior knowledge about the co-occurrence statistics of these signals. This gives rise to a final estimate of the sensory signals  $\hat{S}^{MAP} = (\hat{S}_V^{MAP}, \hat{S}_H^{MAP})$  based on the posterior distribution  $p(S_V, S_H | z_V, z_H) \propto p(z_V, z_H | S_V, S_H)p(S_V, S_H)$ , which balances the benefit of reduced variance with the cost of a potential bias in the estimate (Fig. 12.5, right column). Note that this step does not yet provide an estimate of the world property  $S_W = (S_{W_V}, S_{W_H})$  or the biases  $B = (B_V, B_H)$ . How we estimate  $S_W$  and  $B$  will be discussed in the later section, "From Integration to Remapping." However, from the MAP estimate of the sensory signals we can derive the best estimate of the current discrepancy  $D$  between the signals, which is  $\hat{D}^{MAP} = \hat{S}_V^{MAP} - \hat{S}_H^{MAP}$ .

The posterior estimate  $\hat{S}^{MAP}$  is shifted with respect to the likelihood  $\hat{S}$ . This shift is highlighted by the arrow in Figure 12.5, right column. The length of the arrow indicates the strength of the integration, whereas the direction of the arrow indicates the weighting of the sensory estimates. In the following we will more closely investigate this shift (captured by the two parameters of the arrow) for the three values of  $\sigma_x^2$ .

If the prior is flat ( $\sigma_x^2 \rightarrow \infty$ ; Fig. 12.5, top row), the posterior becomes identical to the likelihood function, which implies that the multisensory estimates are not integrated but kept independent, that is, they are segregated (no shift). Since the signals are independent, any form of integration in this case would only introduce a bias into the final estimates. Given this situation, there can also be no benefit from integration in the form of reduced variance because the signals do not carry redundant information.

In contrast, a prior with  $\sigma_x^2 = 0$  gives rise to a posterior that results in complete fusion (Fig. 12.5, bottom row). As can be observed from the figure, such an impulse prior denotes the existence of only those signals for which  $S_V = S_H$ . Thus, in the case of fusion, the maximum a posteriori (MAP) estimate  $\hat{S}^{MAP}$  coincides with the prior  $p(S_V, S_H)$ . The direction  $\alpha$  in which the estimate is shifted is solely determined by  $\sigma_V^2$  and  $\sigma_H^2$  of the likelihood function (Bresciani et al., 2006):

$$\alpha = \arctan\left(\frac{\sigma_H^2}{\sigma_V^2}\right). \quad (12.8)$$

In this particular case, the MAP estimate maximally benefits from fusion by acquiring the smallest possible variance in the combined estimate. The prior with  $\sigma_x^2 = 0$  applies to a situation with entirely accurate and perfectly redundant signals. Thus, whatever detected discrepancy exists must be a consequence of measurement noise. This case where  $\sigma_x^2 = 0$  is identical to the previously discussed *standard model* of cue integration, which also assumed unbiased (accurate) signals and estimates derived from the same world property (see section on "Multisensory Integration" in this chapter and Chapter 1).

For cases where  $0 < \sigma_x^2 < \infty$ , the MAP estimate  $\hat{S}^{MAP} = (\hat{S}_V^{MAP}, \hat{S}_H^{MAP})$  is situated midway between the maximum-likelihood estimates  $(\hat{S}_V, \hat{S}_H)$  and the diagonal (Fig. 12.5, middle row). In other words, the result here lies between the "no fusion" case and "complete fusion" case, and thus we refer to it as "partial fusion." The strength of integration is indicated by the length  $L$  of the arrow, which has been normalized to the size of the conflict and can be described as a weighting function between the likelihood and the prior in the direction of  $\alpha$  (the direction of bias  $\alpha$  can be determined from Eq. 12.8):

$$L = \frac{\sigma_{likelihood}^2(\alpha)}{\sigma_{likelihood}^2(\alpha) + \sigma_{prior}^2(\alpha)}. \quad (12.9)$$

Any measured discrepancy  $\hat{D} = \hat{S}_V - \hat{S}_H$  is the result of both measurement noise ( $\sigma_V$  and  $\sigma_H$ ) and an actual discrepancy ( $D = S_V - S_H$ ) due to

bias  $B$  (inaccuracy in  $S_V$  and/or  $S_H$ , assuming  $S_{W_V} = S_{W_H}$ ). Combining the likelihood with the prior, resulting in this weighting function (Eq. 12.9), provides the best balance between reliability and accuracy of the estimates of the sensory signals (in the MAP sense). The overall variance of the final estimate resulting from partial integration of the sensory signals lies in between that resulting from pure segregation and complete integration (Fig. 12.5, right column). It must be noted, however, that the final estimate can only profit from the integration process to the extent to which the signals are redundant. Thus, this weighting scheme constitutes the best balance between the costs of introducing a bias in the estimates and benefits of reducing their variances. The remaining difference in the MAP estimates  $\hat{D}^{MAP} = \hat{S}_V^{MAP} - \hat{S}_H^{MAP}$  corresponds to the best current estimate of the actual discrepancy  $D$  between the sensory signals. The predictions of this model, both regarding bias and variance, have been confirmed by an experimental study of the perceived quantity of visual and haptic events (Bresciani et al., 2006).

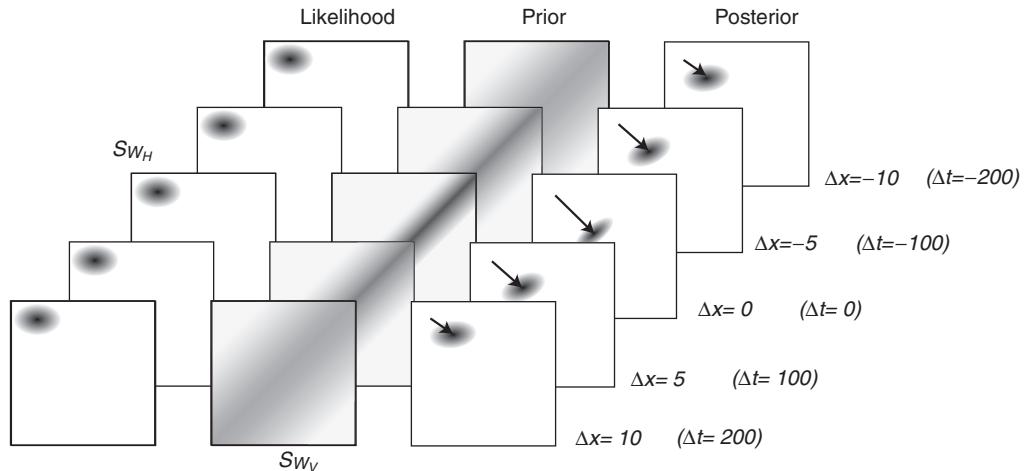
This theoretical framework can also explain how we can learn to integrate over an artificially enforced, statistical relationship between two arbitrary signals (Ernst, 2007). In this study, participants were trained by presenting previously unrelated aspects of a stimulus, for example, the luminance and the stiffness of an object, in correlation for some time. Participants learned this correlation; they began to exhibit integration of the two aspects of the stimulus which were previously unrelated. This was interpreted as the learning of a new prior probability that certain combinations of the two stimulus aspects—luminance and stiffness—are likely to co-occur. Once such a relationship is learned, the newly acquired prior knowledge can be used to integrate the estimates and therefore observers can benefit from a reduction in estimation noise. Thus, during the experiment, the participants switched their behavior from treating the estimates as completely independent to a more intermediate perception of the estimates exhibiting “partial fusion.”

## BREAKDOWN OF INTEGRATION

It is important to note that in the model described in the previous section, the extent of the discrepancy between the maximum-likelihood estimates  $\hat{D} = \hat{S}_V - \hat{S}_H$  does not influence the integration process (i.e., whether estimates are integrated or segregated). The weighting between the estimates, that is, the weighting between the likelihood and the prior, as well as the direction of shift  $\alpha$ , are all independent of the extent of the discrepancy given the assumptions of this model. Thus, this model so far does not capture the breakdown of integration. This is because the shape of the prior and the shape of the likelihood are both assumed to be Gaussian. The problem arises at larger conflicts between signals where, in order to behave robustly, integration should break down. Roach, Heron, and McGraw (2006) suggested relaxing the Gaussian assumption to account for this possibility. In particular, they introduced “heavy tails” to the Gaussian distribution of the prior. This transforms the prior in a very sensible way: Close to the diagonal the prior by and large keeps its Gaussian shape with a reasonable variance. Far from the diagonal the prior does not approach zero probability as a Gaussian would, but maintains a nonzero probability. In essence Roach et al. (2006) suggest a linear combination of a flat coupling prior that is used for modeling segregation (Fig. 12.5, upper row) and a coupling prior that is used for modeling partial fusion or fusion. As a result, the system continues to behave as it did without the long tails when the discrepancies are reasonably small, since the central Gaussian part of the prior plays the dominant role. For larger discrepancies, however, this prior ensures that the process converges toward segregation, because of the increased influence of the flat part of the prior.

This model can further be extended to *orthogonal dimensions* to include, for example, spatial and temporal discordance as well.<sup>5</sup>

<sup>5</sup>We call a conflict along the dimension to be estimated (e.g., size) discrepancy, whereas when we refer to a conflict in an orthogonal dimension (e.g., space or time) we call this discordance.



**Figure 12.6** Schematic illustration demonstrating robust estimation, that is, the breakdown of integration. The coupling prior is assumed to be of Gaussian shape with heavy tails (Roach et al., 2006). The variance of the Gaussian increases with increasing spatial or temporal discordance between the two signals, reflecting a lower correlation between the signals (Fig. 12.3B). Thus, with small discrepancies between the  $S_V$  and  $S_H$  signals, the weight of the prior decreases with temporal asynchrony and spatial disparity and so the effect of integration disappears. With large spatial inconsistencies or temporal asynchronies the two signals can then be perceived independently of one another as the correlation tends to disappear and the coupling prior becomes flat. (Adapted with permission from Ernst, 2007. Copyright ARVO.)

The conceptual basis of the model is illustrated in Figure 12.3B. We assume that under most circumstances, objects and the environment tend to change in their properties over space and time in a smooth, continuous way rather than a discontinuous and chaotic manner. Thus, despite the spatial or temporal discordance, generally there will still be a correlation between the multisensory signals. This correlation implies that despite some spatial and temporal discordance there is still redundancy in the multisensory signals. This redundancy should be used by the brain to improve its estimates. An example of a distribution of spatially discordant  $S_V$  and  $S_H$  size signals is indicated in the lower panel of Figure 12.3B. With increasing spatial discordance  $\Delta x$ , this correlation becomes weaker and weaker until finally the co-occurrence statistics of signals derived from vision and touch will result in a flat distribution. This change in the co-occurrence statistics with increasing spatial or temporal discordance is illustrated in Figure 12.6. The left

column shows a likelihood function, which is identical for all five situations depicted since the sensory measurements are assumed to be identical in all cases. The effect of spatial and temporal discordance is reflected in the prior. For  $\Delta x = 0$  ( $\Delta t = 0$ ) such a prior would resemble a central Gaussian with intermediate variance (analogous to Fig. 12.5, middle row), which also has heavy tails to account for the integration breakdown with increasing disparity in the size estimates (the flat tails are indicated by the gray background in the prior). With increasing spatial or temporal discordance ( $\Delta x \neq 0$  or  $\Delta t \neq 0$ ), the variance of the central part of the prior increases. This is because the prior probability of encountering combinations of  $S_V$  and  $S_H$  signals, for which the discrepancy  $D = S_V - S_H$  is large, increases with the discordance in space or time ( $\Delta x$  or  $\Delta t$ ). As a consequence, as the discordance in space or time increases, the influence that the Gaussian part of the prior exerts on the likelihood function decreases. This process is represented

by the arrows in Figure 12.6 (right column). This phenomenon corresponds to a breakdown of the integration process across space and time, which upon experimentation manifests itself as the spatial and temporal windows of integration.

The exact shape of the prior distribution reflects the co-occurrence statistics of the sensory signal values  $S_V$  and  $S_H$ . This in turn determines the point at which the integration falloff occurs and therefore also determines the dimensions of the temporal or spatial window of integration. It is likely that all these priors have flat tails, because even at large discrepancies there will always be some remaining probability of encountering outliers in the co-occurrence statistics. The tails enable the independent treatment of signals at large inconsistencies. In principle, it should be possible to reconstruct the observers' embodiment of such a prior from experiments that measure the spatial and temporal integration windows. This could be achieved, for example, by extending the methods introduced by Stocker and Simoncelli (2006) to this two-dimensional estimation problem.

Recently, a few other approaches have been proposed to model this elusive aspect of the robustness or breakdown of integration. Some of these methods have also been described in this book (Chapters 2 and 13). We will discuss two of the more prominent proposals in this direction, both of which closely resemble the proposal presented here. The first proposes that the likelihood function is a mixture of Gaussians (Knill, 2007b) to explain the breakdown of integration, whereas the second approach formalizes the concept of causal inference to achieve the same purpose (Körding et al., 2007). Both approaches model the transition from fusion to segregation successfully; however, they both relate to special cases and specific scenarios for which they might be considered optimal.

The mixture-of-Gaussians approach by Knill (2007b) refers to a specific scenario in which a texture signal to slant is modeled by a likelihood function, which is composed of a central Gaussian with heavy tails. This proposal resembles what we have discussed earlier in this chapter, except that the heavy tails are added

to the probability distributions of one of the sensory estimates and not to a coupling prior. The primary argument in this theory is that in order for texture to be a useful signal, we must make some prior assumptions about the isotropy of the texture that, in statistical considerations, could possibly fail in some cases. This argument provides a suitable justification for the use of heavy tails. The argument, however, is specific to the texture signal and can therefore not be easily extended to other within- or cross-modal sensory signals.

The second proposal attempts to formalize the concept of causal inference to model why integration breaks down with highly discrepant information (Körding et al., 2007). The proposal has the same intuitive basis that we have been referring to from time to time, that is, segregation at large discrepancies, integration when there is no apparent discrepancy. This model, however, concentrates on the causal attribution aspects of combining different signals. Two signals could either have one common cause, if they are generated by the same object/event, or they may have different causes when generated by different objects/events. In the former case, the signals should be integrated, and in the latter case they should be kept apart. The model takes into account a prior probability  $p_{common}$  of whether a common source or separate sources exist for a given set of multisensory signals.  $p_{common} = 1$  corresponds to perfect knowledge that there is a common source and thus complete fusion. As discussed previously, complete fusion can be described by a coupling prior corresponding to an impulse prior with  $\sigma_x^2 = 0$ .  $p_{common} = 0$  corresponds to complete knowledge that there are two independent sources and thus complete segregation. Complete segregation was previously described by a flat coupling prior with  $\sigma_x^2 \rightarrow \infty$ . Whenever two sensory signals are detected, in general there will be some probability  $p_{common}$  of a common cause and some probability  $1 - p_{common}$  of independent causes. This probability depends on many factors such as, for example, temporal delays, visual experience, context, and many more (Körding et al., 2007), so it is not easy to predict. In any case, however, it will lead to a weighted

combination of the two priors for complete fusion and segregation, and will thus in essence be analogous to a coupling prior, which has the form of an impulse prior with flat, heavy tails (Körding et al., 2007, supplement). In this sense, the causal-inference model is a special case of the model described earlier. It does not allow for variance in the prior describing the common cause (i.e., the impulse prior), because just like the standard model of integration (see Chapter 1), the causal-inference model is based on the assumption that all sensory signals with a common cause are perfectly correlated and accurate (i.e., the sensory estimates are assumed to be unbiased). Because it does not consider a weaker correlation between the co-occurring signals (i.e., the situation illustrated in Fig. 12.3B) and because it does not take into account the (in)accuracy of the signals (i.e., the situation in Fig. 12.3A), the causal-inference model does not optimally balance the benefits and costs of multisensory integration, that is, reduced variance and potential biases, respectively.

## REMAPPING

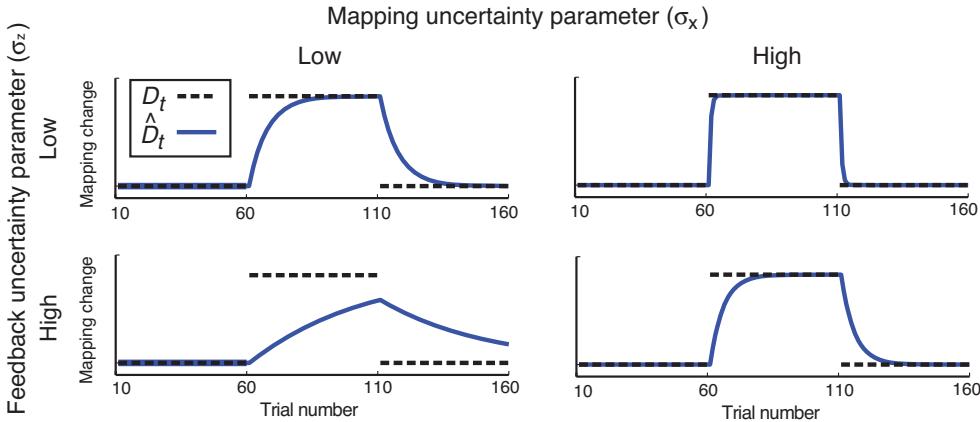
As discussed earlier, multisensory integration breaks down with increasing discrepancy between the estimates. However, if the discrepancy is systematic and persists over several measurements, we adapt to such a discrepancy and doing so brings the conflicting sensory maps (or sensorimotor maps) back into correspondence. This process of adaptation is therefore also referred to as remapping or recalibration. In this section, we review optimal linear models of remapping in the context of a visuomotor task. In the next section, we apply this model to the problem of combining visual and haptic size signals while simultaneously determining the best remapping of each.

There are many examples of such sensory and sensorimotor adaptation processes (e.g., Adams, Banks, & van Ee, 2001; Bedford, 1993; Frissen, Vroomen, & de Gelder, 2003; Pick et al., 1969; Welch, 1978; Welch & Warren, 1980, 1986). The most classic examples of this phenomenon are the experiments on prism shifts first studied by Hermann von Helmholtz (1867). In these

experiments observers were asked to point to a visual target in “open loop” using fast pointing movements. Here, the use of open loop refers to the absence of online feedback to control the movement. The visual feedback can only be procured at the end of the pointing movement upon observing the location where the finger landed. Let this position of the estimated location of the feedback signal be  $\hat{S}_F$  and the estimated target location be  $\hat{S}_L$ . After each trial of such open loop pointing, an error in pointing response can be detected that corresponds to the difference between the feedback- and target-position estimates:  $\hat{D} = \hat{S}_F - \hat{S}_L$ . It is this error that adaptation seeks to minimize.

A typical visuomotor adaptation experiment consists of three phases: a baseline, in which the accuracy of pointing performance is assessed (Fig. 12.7, trial <60). Once the baseline is established, observers receive spectacles fitted with prisms that shift the visual world by some constant amount (e.g., 10°). Once observers wear the prism-fitted spectacles, they exhibit an initial error in their pointing response, which is equivalent to the extent of the prism shift. After only a few pointing movements, however, observers begin to correct for the error induced by the prism and eventually “adapt” to this change (Fig. 12.7, 60 ≤ trial ≤ 110). After adaptation has been achieved, the removal of these prism glasses results in recalibration back to baseline (Fig. 12.7, trial >110).

An interesting aspect of this phenomenon is the rate at which people adapt to these changes. This rate varies enormously depending on the experimental condition. For instance, the rate of adaptation strongly depends on the nature of the conflicting signals provided to the observer. In visuomotor tasks, like pointing to targets, usually the first few trials after wearing prism-spectacles are sufficient for reaching an almost constant minimization of the error, that is, reaching an asymptote for the newly introduced change. In contrast, adaptation purely within the visual domain, for instance, for texture and binocular disparity signals, has been known to take up to several days until adaptation saturates and a constant minimization of the error has been achieved (Adams et al., 2001). Four



**Figure 12.7** Kalman-filter responses to step changes. The dashed black lines in each panel represent the relationship between the position of the reach endpoint and the position of the visual feedback. This relationship is the visuomotor mapping. As in our experiments, there are three phases: prestep (trials 1–60), step (61–110), and poststep (111–160). A first step change in the mapping occurs at the end of the prestep phase; the initial mapping is then restored after the step phase. The blue curves represent the visuomotor mapping estimates  $\hat{D}_{t+}^{\text{MAP}}$  over time. The upper and lower rows show models of the estimates when the measurement uncertainty  $\sigma_z^2$  is small and large, respectively. An increase in  $\sigma_z^2$  causes a decrease in adaptation rate. The left and right columns show responses when the mapping uncertainty  $\sigma_x^2$  is small and large, respectively. An increase in  $\sigma_x^2$  causes an increase in adaptation rate; the effect is larger when  $\sigma_z^2$  is large. (Adapted with permission from Burge et al., 2008. Copyright ARVO.)

examples of adaptation profiles with different rate parameters are provided in Figure 12.7. Even though adaptation has been actively researched for over a 100 years, the search for a computational framework for it only began in recent times with models that tried to describe the process underlying remapping (e.g., Baddley, Ingram, & Miall, 2003; Burge, Ernst, & Banks, 2008; Ghahramani et al., 1997).

In 2008, we investigated how the statistics of the environment and the system together influence the rate of adaptation in visuomotor tasks (Burge et al., 2008). The problem can be formulated in a manner almost analogous to that faced in integration. When a conflict  $\hat{D}_t = \hat{S}_{F,t} - \hat{S}_{L,t}$  is detected on a given trial  $t$ , which in this case would be the difference between the estimated feedback and target positions, the perceptual system must ask itself, what is the source of this conflict.<sup>6</sup> Upon consideration, we find that the

answer is twofold: The conflict could be caused by an actual discrepancy between the sensory (or the sensorimotor) maps  $D_t$ . Alternatively, it could merely be due to measurement noise  $\sigma_z^2$  when acquiring the sensory estimates  $\hat{D}_t$ . If this latter is indeed the case and the discrepancy is caused solely by measurement noise, there would be a new random discrepancy from trial to trial, which would best be ignored by the system. In other words, the system should not attempt to adapt to this randomly fluctuating change in discrepancy caused by measurement noise because to do so would actually make things worse. In sharp contrast, if the discrepancy instead arose due to an actual mismatch in the sensorimotor maps  $D_t$ , it would cause a systematic and sustained discrepancy over trials. Because the occurrence of this discrepancy is persistent and systematic, it would be appropriate for the system to adapt to it.

Analogous to what has been discussed for integration, also for remapping the estimates of both types of error, random versus systematic, contain uncertainty. That is, on a given trial the

<sup>6</sup>In the earlier example, we would define  $\hat{D}_t = \hat{S}_{V,t} - \hat{S}_{H,t}$ .

system can only determine the discrepancy with some uncertainty. The measure of uncertainty for random errors is the variance  $\sigma_z^2$  of the measurement  $z$ . As noted in the previous sections on integration, detecting a systematic error presents more challenge for the system. This is because such an error cannot be determined from one trial observation alone. We must accumulate prior knowledge about the error signal over several observations and use this information to successfully identify a systematic error. Those data, however, also contain uncertainty: the uncertainty  $\sigma_x^2$  associated with the mapping.

Since it is likely that visuomotor tasks contain both systematic and random errors, the nervous system must be able to weight the error estimates flexibly based on their relative uncertainties to solve this credit-assignment problem and to create an optimal estimate of the current mapping. We now turn to a computational framework that formalizes these arguments.

Let us consider that the purpose of the system is primarily to obtain the best possible estimate of the visuomotor mapping in order to remain accurate. The best estimate of the current systematic discrepancy on a given trial,  $\hat{D}_{t+}^{MAP}$  (the MAP estimate derived from the posterior), is a weighted average of the conflict currently measured,  $\hat{D}_t = \hat{S}_{F,t} - \hat{S}_{L,t}$  (the MLE estimate), and the prediction based on past history,  $\hat{D}_{t-}$  (derived from the prior):

$$\begin{aligned}\hat{D}_{t+}^{MAP} &= w_x \hat{D}_{t-} + w_z \hat{D}_t \\ &= \hat{D}_{t-} + K(\hat{D}_t - \hat{D}_{t-}).\end{aligned}\quad (12.10)$$

The value  $K$  is a proportion of the error signal by which the visuomotor mapping is adjusted. In the framework we propose further, we refer to this proportion as the Kalman gain. The “+” on the index indicates that this conflict estimate is used in the next trial to update the mapping; the “−” on the index indicates that this prior information is derived from previous trials, whereas no modifier on the index indicates that it is the measurement derived on the current trial. In an optimal scenario, the weights would be inversely proportional to the relative uncertainties associated with error

estimates based on measurements and prior knowledge:

$$w_x = \frac{\sigma_z^2}{\sigma_z^2 + \sigma_x^2} \quad \text{and} \quad w_z = \frac{\sigma_x^2}{\sigma_z^2 + \sigma_x^2}. \quad (12.11)$$

From Eqs. 12.10 and 12.11, we obtain

$$K = \frac{\sigma_x^2}{\sigma_z^2 + \sigma_x^2}. \quad (12.12)$$

Since  $\hat{D}_{t+}^{MAP}$  is the optimal current estimate of the systematic error that determines the discrepancy, recalibration in any given trial should occur based on this combined estimate.

Adaptation is an iterative process where every trial  $t$  results in an updated combined estimate of the current error signal, which is used for updating the prior in the next step, thereby enabling the efficient tracking of the changes that occur in the mapping. Many experiments show that the brain can adapt under quite complex conditions. For the sake of simplicity, however, here we consider a linear system, which has achieved steady state. Under these assumptions, and following our arguments for Bayesian optimality, the Kalman filter presents an optimal solution to these modeling efforts (for the derivation, refer to Burge et al., 2008). In doing so, we treat the performance of a visuomotor task as a control system in which the error signal is adjusted by the proportion  $K$ , which represents the Kalman gain of such a system.

Figure 12.7 shows the response of a Kalman-filter model to step changes in the mapping. Such a step change is analogous to introducing a prism and later removing it. As the filter adjusts the visuomotor mapping, the error between target and reach position decreases exponentially with time. In other words, human subjects compensate for the error on a trial-by-trial basis to achieve exponentially a constant asymptote at which they have minimized their error. Therefore, we use the exponent  $\lambda$  to express the adaptation rate, which is a function of  $K$ :

$$\lambda = -\log(1 - K). \quad (12.13)$$

From this equation, we find that the model predicts faster adaptation rates for higher gains and low adaptation rates for lower gains. The measurement uncertainty  $\sigma_z^2$  and the mapping uncertainty  $\sigma_x^2$  affect the Kalman gain and thus the adaptation rate in contrasting ways (Eq. 12.12). These opposing effects are illustrated in Figure 12.7. With an increase in measurement uncertainty  $\sigma_z^2$ , the adaptation rate slows down, whereas with an increase in mapping uncertainty  $\sigma_x^2$ , adaptation becomes faster.

These predictions have been tested empirically by systematically varying the measurement noise using various blur conditions on the visual feedback signals, thus making them less reliable to estimate (Burge et al., 2008). They found that observers did indeed adapt more slowly with an increase in the blur of the feedback stimuli. When they introduced a perturbation into the mapping on a trial-by-trial basis instead of blurring the feedback signal, however, they found that a random but statistically stationary error in the feedback did not elicit any change in the rate of adaptation. That trial-by-trial variation did not affect the rate of recalibration suggests that the measurement noise may be estimated online in any given trial, but not over trials.

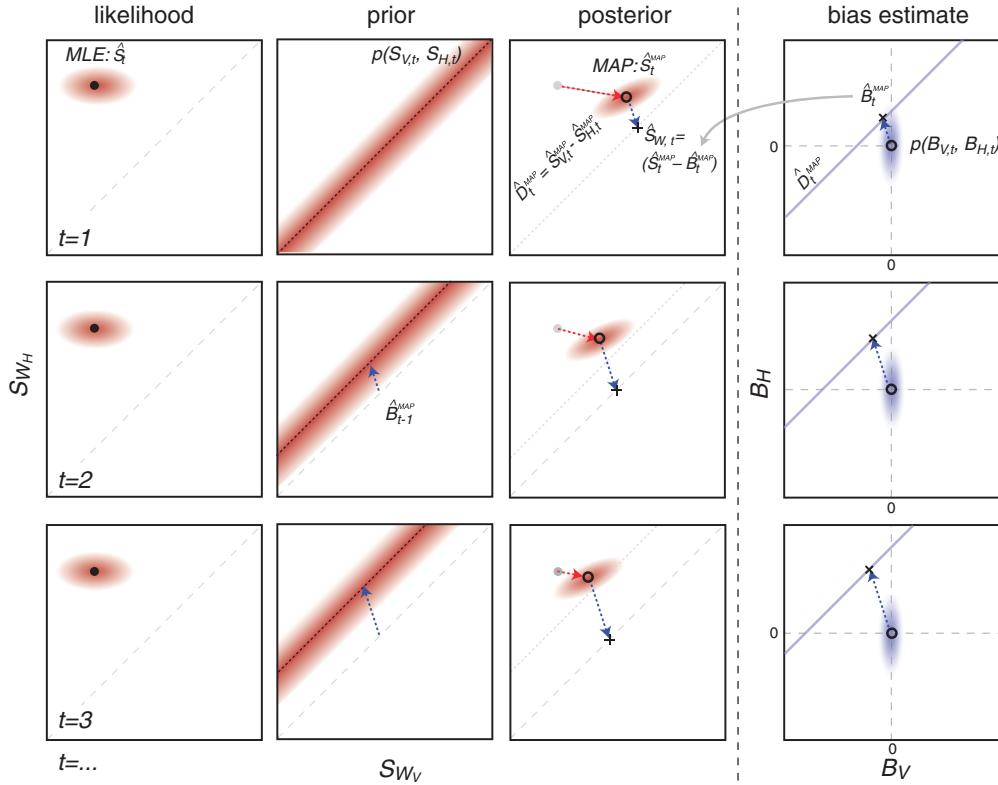
In a second experiment Burge et al. (2008) perturbed the mapping from trial to trial with time-correlated noise in a random-walk fashion. To put it simply, in each trial a new random variable drawn from a Gaussian distribution was added to the previous mapping. If correctly learned, this manipulation affects the mapping uncertainty as the mapping is constantly changing in a time-correlated fashion. Consistent with the predictions of the optimal adaptor, the results showed an increased adaptation rate for an increase in the variance of the random-walk distribution. In conclusion, it seems that to a first approximation (e.g., assuming stationary statistics) the Kalman-filter model is a good predictor of human adaptation performance.

## FROM INTEGRATION TO REMAPPING

In this section we apply the Bayesian (Kalman-filter) model of remapping to the visual-haptic

size-estimation task and combine it with the partial-integration model from earlier in this chapter. This is illustrated in Figure 12.8. We assume there is a sequence of trials at times  $t$  in which the observer has sensory estimates  $\hat{S}_{V,t}$  and  $\hat{S}_{H,t}$ , and tries to estimate  $S_W = (S_{W_V}, S_{W_H})$ . For simplicity, we assume the perceptual situation to be constant throughout the trials, so that  $S_W$ ,  $S_i$ , and  $B_i$  are all independent of  $t$ . Furthermore, for now we assume that  $S_{W_V} = S_{W_H}$ , which implies that we are estimating the identical world property by vision and touch. The initial situation for estimating  $S_W$  is that there may exist an unknown additive bias  $B = (B_V, B_H)$  in the visual and haptic signals  $S = (S_V, S_H) = (S_{W_V} + B_V, S_{W_H} + B_H)$  leading to a discrepancy  $D = S_V - S_H$  between the sensory signals. At first, we do not know these biases, so at time step  $t = 0$ , before any measurement is performed, the initial bias estimate is  $\hat{B}_0 = 0$ , the initial prediction for the discrepancy is  $\hat{D}_{t=0} = 0$ , and the initial coupling prior  $p_0(S_V, S_H) = p(S_{V,0}, S_{H,0})$  is unbiased, that is, it is centered on the diagonal  $S_V = S_H$ . For every time step  $t = 1, 2, 3, \dots$ , the observer begins by deriving the maximum-likelihood estimate  $\hat{S}_t = (\hat{S}_{V,t}, \hat{S}_{H,t})$  of the current signals  $S_t = (S_{V,t}, S_{H,t})$ . These MLE estimates contain a discrepancy  $\hat{D}_t = \hat{S}_{V,t} - \hat{S}_{H,t}$ . In the leftmost column of Figure 12.8 this is indicated by the red Gaussian blobs being off the diagonal (equivalent to Fig. 12.5, left column).<sup>7</sup> The variance of the likelihood function  $\Sigma_z$  indicates the measurement uncertainty. Next, to solve the credit-assignment problem of whether this discrepancy  $\hat{D}_t$  is caused by noise  $\Sigma_z$  or an actual difference  $D$  between the signals, the Bayesian integration scheme is applied, combining the maximum-likelihood estimate with prior knowledge about the joint distribution of  $S_{V,t}$  and  $S_{H,t}$ , that is, the mapping between the signals. That is, the column labeled “prior” in Figure 12.8 shows an example of an intermediate “coupling prior” with variance  $\sigma_x^2$ . This variance

<sup>7</sup>For illustrative purposes we assume that at every trial the same noise value  $\varepsilon$  is added to the measurement  $z_{i,t} = S_{i,t} + \varepsilon_{i,t}$  so the likelihood function is identical in each row of Figure 12.8.



**Figure 12.8** Illustration of the link between integration and remapping of visual and haptic size estimates. The leftmost column illustrates the maximum-likelihood estimates  $\hat{S}_t = (\hat{S}_{V,t}, \hat{S}_{H,t})$  indicated by a dot with the corresponding measurement noise  $\Sigma_z$  indicated by the red Gaussian blob. The column labeled “prior” gives a coupling prior  $p(S_{V,t}, S_{H,t}) = p_0(S_{V,t} - \hat{B}_{V,t-1}^{MAP}, S_{H,t} - \hat{B}_{H,t-1}^{MAP})$  with corresponding mapping uncertainty  $\sigma_x^2$  indicated by the red shaded area. The column labeled “posterior” shows the maximum a posteriori (MAP) estimate, indicated by the  $\bullet$ , together with its variance. The MAP estimate is the result of the Bayes’ product between likelihood and prior. The amount of integration and the weighting of the signals are given by the length and the orientation of the red arrow, respectively, just as in Figure 12.5. The estimate of discrepancy resulting from the MAP estimate is given by  $\hat{D}_{t+}^{MAP} = \hat{S}_{V,t}^{MAP} - \hat{S}_{H,t}^{MAP}$ . To determine the part of  $\hat{D}_{t+}^{MAP}$  that can be attributed to a visual or haptic bias is again an ambiguous problem. This new credit-assignment problem is solved in the rightmost column labeled “bias estimate.” Here the ambiguous  $\hat{D}_{t+}^{MAP}$  estimate is represented by the diagonal line. Additionally, there is prior information  $p(B_V, B_H)$  about potential biases occurring in the visual and haptic modality, which is indicated by the blue Gaussian blob. The discrepancy estimate combined with the bias prior according to Bayes’ rule results in the current bias estimate  $\hat{B}_t^{MAP}$ . This resulting bias estimate is used for shifting the coupling prior in the next time step. The estimate  $\hat{B}_t^{MAP}$  is indicated by  $x$  and the blue arrow. The size estimate of the object is the combination of the MAP and the bias estimate according to  $\hat{S}_{W,t} = \hat{S}_t^{MAP} - \hat{B}_t^{MAP} = (\hat{S}_{V,t}^{MAP} - \hat{B}_{V,t}^{MAP}, \hat{S}_{H,t}^{MAP} - \hat{B}_{H,t}^{MAP})$ . This is indicated by the sum of the red and blue arrow in the “posterior” column. Each row provides a new time step in the remapping process. Repeating the same estimation over several trials  $t$ , the bias estimate  $\hat{B}_t^{MAP}$ , as indicated by the blue arrow, is exponentially increasing so that in the end the system reaches the calibrated steady state.

corresponds to the mapping uncertainty. Applying Bayes' rule  $p(S_{V,t}, S_{H,t} | z_{V,t}, z_{H,t}) \propto p(z_{V,t}, z_{H,t} | S_{V,t}, S_{H,t})p(S_{V,t}, S_{H,t})$  results in the optimal current estimates of the sensory signals  $\hat{S}_t^{MAP} = (\hat{S}_{V,t}^{MAP}, \hat{S}_{H,t}^{MAP})$ , thereby maximally reducing the variance in the sensory estimates while at the same time providing the best possible estimate of the current discrepancy  $\hat{D}_{t+}^{MAP} = \hat{S}_{H,t}^{MAP} - \hat{S}_{V,t}^{MAP}$  at time step  $t$ . Thus, the MAP estimate of the discrepancy  $\hat{D}_{t+}^{MAP}$  is smaller than  $\hat{D}_t$  to the extent that the two sensory signals are coupled. The result of combining likelihood with prior knowledge using Bayes' rule is illustrated in Figure 12.8 in the column labeled "posterior." The result of integration corresponds to Eq. 12.10:  $\hat{D}_{t+}^{MAP} = w_x \hat{D}_{t-} + w_z \hat{D}_t = \hat{D}_{t-} + K(\hat{D}_t - \hat{D}_{t-})$ . This integration process, illustrated by the red distributions and the red arrow, is identical to what was shown in Figure 12.5. The MAP estimate  $\hat{S}_t^{MAP} = (\hat{S}_{V,t}^{MAP}, \hat{S}_{H,t}^{MAP})$  at each time step is the best current estimate of the size signals available. The best current discrepancy estimate between the size signals corresponds to  $\hat{D}_{t+}^{MAP} = \hat{S}_{V,t}^{MAP} - \hat{S}_{H,t}^{MAP}$ .

Note, up to now we have no estimate of bias  $\hat{B}_t$  and no estimate of visual and haptic object size  $S_W = (S_{W_V}, S_{W_H})$ . What we do have is the discrepancy estimate  $\hat{D}_{t+}^{MAP}$ , but to what extent the visual and haptic biases contribute this discrepancy

$$\begin{aligned}\hat{D}_{t+}^{MAP} &= (S_{W_V} + \hat{B}_{V,t}^{MAP}) - (S_{W_H} + \hat{B}_{H,t}^{MAP}) \\ &= \hat{B}_{V,t}^{MAP} - \hat{B}_{H,t}^{MAP}, \text{ given that we are} \\ &\quad \text{assuming } S_{W_V} = S_{W_H},\end{aligned}\quad (12.14)$$

is still unknown. This ambiguity in the discrepancy estimate after integration is indicated by the blue diagonal line in the rightmost column of Figure 12.8. It illustrates that there is an infinite combination of visual and haptic biases that are consistent with  $\hat{D}_{t+}^{MAP}$ . For now, we assume that we know for sure that the visual and haptic sizes are identical ( $S_{W_V} = S_{W_H}$ ), so the discrepancy estimate given by the blue line contains no noise, that is, is not blurry. The attribution of visual and haptic bias to the discrepancy estimate is a second credit-assignment problem, and in order to solve it we need additional prior knowledge.

In the following we will discuss how to best resolve this new credit-assignment problem. Ghahramani and colleagues (1997) proposed that the discrepancy in the sensory estimates should be resolved in proportion to their variances ( $\sigma_V^2, \sigma_H^2$ ), that is, more credit should be given to a signal with higher variance. However, since the variance of an estimate does not necessarily determine the probability of it containing a bias (i.e., its contribution to the discrepancy), this might lead to a suboptimal strategy. A better way to resolve the credit assignment problem resulting from the "bias ambiguity" may be to use prior knowledge about the probability of the signals being biased  $p(B_V, B_H)$ . We call this the "bias prior." We need to use prior knowledge because there is no direct information in the sensory signals about whether they are accurate or biased. For example, if the estimates derived from the haptic modality have often been biased in the past, it is more likely that the haptic modality provides the biased signal also in the current situation. This prior knowledge encoding the probability of a bias in a sensory signal is indicated by the blue Gaussian blob in the rightmost column of Figure 12.8. The variance of this prior distribution determines the probability of the signal to be biased. In the example of Figure 12.8, the visual signal is less likely to be biased than the haptic signal. Consequently, in the absence of any other evidence as to what may have caused the discrepancy, the ambiguity in the discrepancy estimate will be resolved once again using Bayes' rule. This time we use Bayes' rule to combine the discrepancy estimate  $\hat{D}_{t+}^{MAP}$  with this bias prior  $p(B_V, B_H)$ . This will result in the current best bias estimate  $\hat{B}_t^{MAP} = (\hat{B}_{V,t}^{MAP}, \hat{B}_{H,t}^{MAP})$  indicated by the blue arrow in the rightmost column of Figure 12.8. The proportion  $\hat{B}_{V,t}^{MAP}/\hat{B}_{H,t}^{MAP}$  and thus the direction of the blue arrow are solely dependent on the variance of the bias prior  $p(B_V, B_H)$ . Now that we have a bias estimate we also have an estimate of the visual and haptic size of the object, which was our objective from the start of this chapter. The visual and haptic sizes are given by  $\hat{S}_{W,t} = \hat{S}_t^{MAP} - \hat{B}_t^{MAP} = (\hat{S}_{V,t}^{MAP} - \hat{B}_{V,t}^{MAP}, \hat{S}_{H,t}^{MAP} - \hat{B}_{H,t}^{MAP})$  as indicated in Figure 12.8 by the

sum of the red and blue arrows in the column labeled “posterior.”

With this our estimation problem is solved and at the end of the time step we have the best current estimate of the sensory signals  $\hat{S}_t^{MAP}$ , the sizes of the objects  $\hat{S}_{W,t}$ , and the biases  $\hat{B}_t^{MAP}$ . However, to achieve even more accurate estimates in the future, we have to recalibrate our system based on these bias estimates.

The iterative recalibration process is described next. Each row in Figure 12.8 denotes a new time step  $t - 1$ . After integration at time step  $t - 1$  the perceptual system is left with a bias estimate  $\hat{B}_{t-1}^{MAP} = (\hat{B}_{V,t-1}^{MAP}, \hat{B}_{H,t-1}^{MAP})$ . It is this bias estimate that is used during recalibration (remapping) to change the mapping at time  $t$  defined by the coupling prior. Thus, the coupling prior at time  $t$  will be shifted to be consistent with the current estimate of the bias, so that  $p(S_{V,t}, S_{H,t}) = p_0(S_{V,t} - \hat{B}_{V,t-1}^{MAP}, S_{H,t} - \hat{B}_{H,t-1}^{MAP})$ , indicated by the blue arrow in the “prior” column of Figure 12.8.<sup>8</sup>

This iterative updating process corresponds to the Kalman-filter approach to remapping that we discussed in the last section. As can be seen from Figure 12.8, while the direction of the blue arrow stays constant, the length of the blue arrow continuously increases with every time step, providing an increasingly accurate estimate of the bias  $\hat{B}_t^{MAP}$  and the world property  $\hat{S}_{W,t}$ . Thereby, it is the discrepancy estimate  $\hat{D}_{t+}^{MAP}$  that determines the extent to which one must adapt at each time step, and this in turn determines the rate of adaptation. This is consistent with the exponential adaptation response discussed in the previous section (Fig. 12.7 and Eq. 12.13). After several time steps the system eventually reaches steady state. This steady state, however, can only be reached if the bias  $B$  is constant over several trials, such as for example when wearing glasses (Fig. 12.3A, third row). In contrast, if the bias  $B$  is constantly changing from trial to trial, such as for example when using different grip postures with every size measurement (Fig. 12.3A, second

<sup>8</sup>The prior update corresponds to a linear shift. In general, however, also the variance of the prior should be updated. This however, goes beyond the scope of this chapter.

row), recalibration will never saturate and reach steady state.

There is empirical evidence that the bias assignment during recalibration can be situation dependent, indicating the use of different bias priors depending on the exact stimulation condition. This was shown by Di Luca, Machulla, and Ernst (2009) in a recent study on visual-auditory adaptation to temporal asynchrony. For colocated visual-auditory signals, they found that it is the visual estimates that exhibit adaptation. On the other hand, when auditory signals were presented via headphones, they found that participants adapted the auditory estimates. In this study they argue that the manipulation, headphones versus no headphones, alters the probability of the auditory estimate to be biased, which is reflected in a change of the bias prior. They argue that this is because, when presented via headphones, the auditory signal becomes the “odd one out” and is thus more likely biased. The change in bias prior with spatial discrepancy (i.e., with headphones) is similar to the change in the coupling prior with spatial discordance we discussed earlier for modeling the breakdown of integration. With headphones the bias prior will have a larger variance for  $B_A$  than for  $B_V$ , so that with headphones it is predominantly the auditory modality that becomes recalibrated.

So far we have not discussed that the discrepancy estimate  $\hat{D}_{t+}^{MAP}$  may contain uncertainty as well. We assumed that we are seeing and touching the object at the same location so that the size of the object is equal in vision and touch  $S_{W_V} = S_{W_H}$ . In this case any discrepancy detected  $\hat{D}_{t+}^{MAP}$  must be the result of a bias so there is no uncertainty in the discrepancy estimate. However, if we consider the situation depicted in Figure 12.3B when we are seeing and touching the object at discrepant locations,  $S_{W_V}$  may be different from  $S_{W_H}$ . This way the discrepancy  $\hat{D}_{t+}^{MAP}$  could either result from a bias in one of the sensory signals or from an actual difference in the visual and haptic sizes of the object ( $S_{W_V}, S_{W_H}$ ). If there is such an uncertainty in the discrepancy estimate  $\hat{D}_{t+}^{MAP}$ , the blue diagonal line representing this discrepancy in Figure 12.8 becomes blurry as well, with the blur representing the

uncertainty. As a result, the Bayes product of the discrepancy estimate with the bias prior will result in a reduced bias estimate  $\hat{B}_t^{\text{MAP}}$  (compared to the situation when  $\hat{D}_{t+}^{\text{MAP}}$  was certain). Subtracting this reduced bias estimate  $\hat{B}_t^{\text{MAP}} = (\hat{B}_{V,t}^{\text{MAP}}, \hat{B}_{H,t}^{\text{MAP}})$  from the MAP estimate  $\hat{S}_t^{\text{MAP}} = (\hat{S}_{V,t}^{\text{MAP}}, \hat{S}_{H,t}^{\text{MAP}})$  will again result in the best current size estimate (Fig. 12.8 sum of red and blue arrow in column labeled “posterior”). However, with the reduced bias estimate, the size estimates for vision and touch will differ:  $\hat{S}_{W,t} = \hat{S}_t^{\text{MAP}} - \hat{B}_t^{\text{MAP}} = (\hat{S}_{V,t}^{\text{MAP}} - \hat{B}_{V,t}^{\text{MAP}}, \hat{S}_{H,t}^{\text{MAP}} - \hat{B}_{H,t}^{\text{MAP}})$ . Given the situation in Figure 12.3B, this is the optimal estimate of the visual and haptic sizes, their biases, and their signals. As  $\hat{B}_t^{\text{MAP}}$  becomes smaller with increasing uncertainty in  $\hat{D}_{t+}^{\text{MAP}}$  the amount of bias recalibration per time step will decrease accordingly.

This brings us to a question that we invoked in the previous section, which we can now answer more comprehensively. This question pertains to why the rate of adaptation varies so much in differing scenarios. The answer is that this is because the rate is a function of both the measurement noise  $\Sigma_z$  and the mapping uncertainty  $\sigma_x$ . A lower measurement noise  $\Sigma_z$  causes an increased adaptation rate and so does higher mapping uncertainty  $\sigma_x$ . Similarly, we have discussed earlier that the extent of integration increases with an increase in measurement noise  $\Sigma_z$  and a decrease of mapping uncertainty  $\sigma_x$ . This dependency between the rate of adaptation and the degree of integration appears to have empirical support. For instance, texture and binocular disparity signals within the visual domain have been shown to be integrated such that they exhibit almost complete fusion (Hillis et al., 2002). This corresponds well to the fact that adaptation of texture and disparity signals can take up to 2 or 3 days (Adams et al., 2001). In contrast, the extent of integration between visual and haptic signals seems significantly far from complete fusion (Ernst, 2005; Hillis et al., 2002), which corresponds to the relatively fast adaptation rate found in the sensorimotor domain and for visual-haptic estimates. From this we may conclude that the mapping uncertainty ( $\sigma_x$ ; the variance of the coupling prior) is very

low for the texture and disparity signals in the visual domain and it is comparatively high for sensorimotor and visual-haptic signals. This makes sense because the two visual signals, texture and disparity, originate from the same receptors on the retina, so discrepancies between these two signals are not very likely to occur. The absence of discrepancies means there is no need for adaptation, and thus the estimates can be mandatorily fused. For visual-haptic estimates, however, the situation is different, because there are different receptors responsible for the sensory signals. A discrepancy between the senses is generated, for example, every time we use tools. For example, when we grasp an object using a gripper or a big glove, we have to adjust the mapping between the visual and haptic signals to use these tools in open loop. To adjust quickly to such novel situations, errors in the feedback signals have to be detected, which they could not be if the signals were mandatorily fused. Thus, in the case of visual-haptic estimates the mapping uncertainty  $\sigma_x$  is higher than in the case of texture and disparity estimates, leading to a faster adaptation rate and to signals that are not mandatorily fused.

## CONCLUDING REMARKS AND SOME OPEN QUESTIONS

We began this chapter by describing the standard model of integration (see Chapter 1) because it has proven successful in the description of many experimental results. During the course of this chapter, our goal was to argue that the process of integration is more complex than the standard model leads us to believe. The standard model seeks to reduce the noise in perceptual estimates through integration and focuses on what we have earlier referred to as the benefits of integration. However, it assumes perfectly correlated and unbiased (accurate) estimates and therefore disregards the potential cost of integration. The price we pay for such an assumption is that we disregard the fact that the estimates may actually become biased during integration. If we take into account that signals may be inaccurate, by balancing costs and benefits in a probabilistic way as we have

suggested here, we achieve a unified framework that can explain the process of integration as well as the process of remapping.

This framework assumes that for both integration and remapping, there exists a precise representation of the probability distributions that represent the measurement process as well as prior knowledge about the mapping between different sensory signals, that is, their co-occurrence statistics, and about the probability of occurrence of a discrepancy between the sensory signals  $S_i$  and the world property  $S_w$ . This includes knowing not only the mean and the variance of the distributions but also the entire shape, because knowledge of the shape of the probability distributions has proven to be crucial, especially in the efforts to model the breakdown of integration. Up to now, however, our knowledge about how these distributions are represented in the human brain is very limited. There are some indications from computational neuroscience that the utilization of such probability distributions may be implemented in a neuronal population code. Pouget, Deneve, and Duhamel (2002) demonstrated that processing with such population codes can perform integration and remapping of sensory estimates equally well (see also Chapter 21; Ma, Beck, Latham, & Pouget, 2006; Pouget et al., 2002). There is also some recent evidence from monkey physiology that Bayes-optimal integration of visual and vestibular estimates of perceived direction of motion may be conducted in populations of neurons in area MSTd of the monkey brain (see Chapter 16; Gu, Angelaki, & DeAngelis, 2008).

One major problem that we face currently with the Bayesian approach is that the prior distributions, which are used to represent the statistics of the sensory signals derived from the environment, are merely postulated. The reason that they are only postulated is that they are not easily measurable. Future research needs to address how priors can be determined, measured, or manipulated independently. There are several ways to achieve this. One method entails the study of how priors are learned in one context and then to investigate how the learning of these priors is transferred to other situations or tasks. This has successfully been demonstrated by

Adams et al. (2004), who adapted the light-from-above prior in one context and demonstrated transfer of the learning effect to another context and task, thereby undoubtedly demonstrating the updating of a priori knowledge. Alternatively, one can indirectly infer the prior by varying the reliability of the signal estimates using a method proposed by Stocker and Simoncelli (2006). Ideally, such an inferred prior distribution should then be compared to the statistics of the environment as measured independently by some physical measurement device. Recently, this approach has been employed by Banks, Burge, and Held (Chapter 11) for the visual perception of depth at an occluding contour. In their experiments, Burge et al. found a good qualitative agreement between the empirically inferred prior and the physically measured statistics of the environment. Future research must strive to demonstrate such correspondence between postulated priors and the statistics of the environment in order to justify the Bayesian approach to human perception.

## ACKNOWLEDGMENTS

We thank Mike Landy for very helpful comments on earlier drafts of this chapter and Devika Narain for outstanding help in the editing process. This work was supported by the EU grant ImmerSence (IST-2006-027141), the EU grant THE (IST-2009-248587), and the HFSP Research Grant (2006) "Mechanisms of associative learning in human perception."

## REFERENCES

- Adams, W. J., Banks, M. S., & van Ee, R. (2001). Adaptation to three-dimensional distortions in human vision. *Nature Neuroscience*, 4, 1063–1064.
- Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience*, 7, 1057–1058.
- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near optimal crossmodal integration. *Current Biology*, 14, 257–262.
- Baddeley, R. J., Ingram, H. A., & Miall, R. C. (2003). System identification applied to a visuomotor task: Near-optimal human performance in a

- noisy changing task. *Journal of Neuroscience*, 23, 3066–3075.
- Bedford, F. L. (1993). Perceptual learning. In D. Medin (Ed.), *The psychology of learning and motivation* (pp. 1–60). New York, NY: Academic Press.
- Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory–visual discordance. *Perception and Psychophysics*, 29, 578–584.
- 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.
- Bresciani, J.-P., Dammeier, F., & Ernst, M. O. (2008). Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events. *Brain Research Bulletin*, 75, 753–760.
- Bresciani, J.-P., & Ernst, M. O. (2007). Signal reliability modulates auditory-tactile integration for event counting. *NeuroReport*, 18, 1157–1167.
- Bresciani, J.-P., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., & Kheddar, A. (2005). Feeling what you hear: Auditory signals can modulate tactile taps perception. *Experimental Brain Research*, 162, 172–180.
- Brewster, D. (1826). On the optical illusion of the conversion of cameos into intaglios and of intaglios into cameos, with an account of other analogous phenomena. *Edinburgh Journal of Science*, 4, 99–108.
- Burge, J., Ernst, M. O., & Banks, M. S. (2008). The statistical determinants of adaptation rate in human reaching. *Journal of Vision*, 8(4):20, 1–19.
- Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society*, 4(Suppl.), 102–118.
- Di Luca, M., Machulla, T.-K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Crossmodal transfer coincides with a change in perceptual latency. *Journal of Vision*, 9(12):7, 1–16.
- Dror, R., Willsky, A., & Adelson, E. (2004). Statistical characterization of real-world illumination. *Journal of Vision*, 4, 821–837.
- Ernst, M. O. (2005). A Bayesian view on multimodal cue integration. In G. Knoblich, I. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Human body perception from the inside out* (pp. 105–131). New York, NY: 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.
- Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8, 162–169.
- Frissen, I., Vroomen, J., & de Gelder, B. (2003). The aftereffects of ventriloquism: Are they sound-frequency specific? *Acta Psychologica*, 113, 315–327.
- Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *The American Journal of Psychology*, 72, 521–529.
- Gepshtain, S., Burge, J., Banks, M. S., & Ernst, M. O. (2005). Optimal combination of vision and touch has a limited spatial range. *Journal of Vision*, 5, 1013–1023.
- Ghahramani, Z., Wolpert, D. M., & Jordan, M. I. (1997). Computational models of sensorimotor integration. In P. G. Morasso & V. Sanguineti (Eds.), *Self-organization, computational maps and motor control* (pp. 117–147). Amsterdam, Netherlands: Elsevier Press.
- Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, 11, 1201–1210.
- Hartung, B., Schrater, P. R., Bülthoff, H. H., Kersten, D., & Franz, V. H. (2005). Is prior knowledge of object geometry used in visually guided reaching? *Journal of Vision*, 5, 504–514.
- Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research*, 179, 595–606.
- 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.
- Hillis, J. M., Watt, S., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4, 1–24.
- Jack, C. E., Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. *Perceptual and Motor Skills*, 37, 967–979.

- Jackson, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, 5, 52–65.
- Kersten D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 44, 271–304.
- Knill, D. C. (2007a). Learning Bayesian priors for depth perception. *Journal of Vision*, 7(8):13, 1–20.
- Knill, D. C. (2007b). 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.
- Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43, 2539–2558.
- Körding, K. P., Beierholm, U., Ma, W., Quartz, S., Tenenbaum, J., Shams, L., & Sporns, O. (2007). Causal inference in multisensory perception. *PLoS ONE*, 2, e943.
- Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration of sensorimotor learning. *Nature*, 427, 244–247.
- Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A*, 18, 2307–2320.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. J. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35, 389–412.
- Ma, J. W., Beck, J. M., Latham, P. E., & Pouget A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9, 1432–1438.
- Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, 81, B1–B9.
- Mamassian, P., Landy, M., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain* (pp. 13–60). Cambridge, MA: MIT Press.
- Myers, A. K., Cotton, B., & Hilp, H. A. (1981). Matching the rate of concurrent tone bursts and light flashes as a function of flash surround luminance. *Perception and Psychophysics*, 30, 33–38.
- Oruç, I., Maloney, T. M., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, 43, 2451–2468.
- Palmer, S. (1985). The role of symmetry in shape perception. *Acta Psychologica*, 59, 67–90.
- Pick, H., Warren, D., & Hay, J. (1969). Sensory conflict in judgments of spatial direction. *Perception and Psychophysics*, 6, 203–205.
- Pouget, A., Deneve, S., & Duhamel, J. (2002). Opinion: A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, 3, 741–747.
- Radeau, M., & Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs. Thomas (1941) revisited. *Psychological Research*, 49, 17–22.
- Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, 89, 1078–1093.
- Roach, N., Heron, J., & Mcgraw, P. (2006). Resolving multisensory conflict: A strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society of London B: Biological Sciences*, 273, 2159–2168.
- Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14, 147–152.
- Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, 187, 802.
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9, 578–585.
- van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1998). The precision of proprioceptive position sense. *Experimental Brain Research*, 122, 367–377.
- van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1999). Integration of proprioceptive and visual position information: An experimentally supported model. *Journal of Neurophysiology*, 81, 1355–1364.
- van Wassenhove, V., Grant, K., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45, 598–607.
- von Helmholtz, H. (1867). *Handbuch der physiologischen Optik* (Vol. 3). Leipzig, Germany: Leopold Voss.
- 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.

- Warren, D. H., & Cleaves, W. T. (1971). Visual-proprioceptive interaction under large amounts of conflicts. *Journal of Experimental Psychology, 90*, 206–214.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*, 598–604.
- Welch, R. B. (1978). *Perceptual modification: Adaptation to altered sensory environments*. New York, NY: Academic Press.
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 638–667.
- Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 25.1–25.36). New York, NY: J. Wiley & Sons.
- Welch, R., DuttonHurt, L., & Warren, D. (1986). Contributions of audition and vision to temporal rate. *Perception and Psychophysics, 39*, 294–300.
- Witkin, H. A., Wapner, S., & Leventhal, T. J. (1952). Sound localization with conflicting visual and auditory cues. *Journal of Experimental Psychology, 43*, 58–67.
- Young, M., Landy, M., & Maloney, L. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research, 33*, 2685–2696.