

Learning to integrate arbitrary signals from vision and touch

Marc O. Ernst

Max Planck Institute for Biological Cybernetics,
Tübingen, Germany



When different perceptual signals of the same physical property are integrated, for example, an objects' size, which can be seen and felt, they form a more reliable sensory estimate (e.g., M. O. Ernst & M. S. Banks, 2002). This, however, implies that the sensory system already knows which signals belong together and how they relate. In other words, the system has to know the mapping between the signals. In a Bayesian model of cue integration, this prior knowledge can be made explicit. Here, we ask whether such a mapping between two arbitrary sensory signals from vision and touch can be learned from their statistical co-occurrence such that they become integrated. In the Bayesian framework, this means changing the belief about the distribution of the stimuli. To this end, we trained subjects with stimuli that are usually unrelated in the world—the luminance of an object (visual signal) and its stiffness (haptic signal). In the training phase, we then presented subjects with combinations of these two signals, which were artificially correlated, and thus, we introduced a new mapping between them. For example, the stiffer the object, the brighter it was. We measured the influence of learning by comparing discrimination performance before and after training. The prediction is that integration makes discrimination worse for stimuli, which are incongruent with the newly learned mapping, because integration would cause this incongruity to disappear perceptually. The more certain subjects are about the new mapping, the stronger should the influence be on discrimination performance. Thus, learning in this context is about acquiring beliefs. We found a significant change in discrimination performance before and after training when comparing trials with congruent and incongruent stimuli. After training, discrimination thresholds for the incongruent stimuli are increased relative to thresholds for congruent stimuli, suggesting that subjects learned effectively to integrate the two formerly unrelated signals.

Keywords: vision, touch, cue integration, Bayesian models of perception, statistical learning

Citation: Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, 7(5):7, 1–14, <http://journalofvision.org/7/5/7/>, doi:10.1167/7.5.7.

Introduction

Our brain constantly receives sensory information from many different sources and modalities. If corresponding sensory signals are derived from the same object or event, these should be integrated; otherwise, they should be kept separate. For example, when moving your head, visual, vestibular, and proprioceptive signals all give rise to an estimate of the heads' position and orientation. Hence, it would make sense for the sensory system to integrate these signals into a coherent representation of head position and orientation. Another example is depth perception: The distance to an object can be estimated from the disparity signal between the two eyes' images, from perspective distortions in the image, from motion parallax, and from many other signals. Integration of such depth signals has been demonstrated repeatedly (e.g., Bühlhoff & Mallot, 1988; Hillis, Watt, Landy, & Banks, 2004; Howard & Rogers, 2002; Knill & Saunders, 2003; Landy, Maloney, Johnston, & Young, 1995).

Each sensory signal is inherently noisy and so is the estimate of the physical property that the sensory signal represents. The advantage of integrating redundant sources of sensory information is that the variance in the integrated estimate can be reduced relative to the

estimates derived individually (e.g., Clark & Yuille, 1990; Ernst & Banks, 2002; Ernst & Bühlhoff, 2004; Ghahramani, Wolpert, & Jordan, 1997; Jacobs, 2002; Landy et al., 1995). For example, when manipulating objects, an object's size can be judged simultaneously by vision and touch (Ernst & Banks, 2002) and should therefore be integrated by our nervous system to benefit from the redundancy in the sensory information. Compared with the noisy estimate from each modality alone, integrating the information from the two modalities yields a more certain estimate of the objects' size. Such a benefit may reveal itself in a size discrimination task when the reduced variance in the size estimate is reflected in an improved discrimination performance compared to size discrimination based on only one modality. The reliability of a sensory estimate can be defined as its inverse variance $r = \sigma^{-2}$. An optimal method for combining sensory information would maximize the reliability of the final (unbiased) estimate. Recently, several studies indicated that the human brain integrates sensory information in such an optimal way (Alais & Burr, 2004; Ernst & Banks, 2002; Helbig & Ernst, 2007; Hillis et al., 2004; Knill & Saunders, 2003; Landy & Kojima, 2001).

However, integration also comes at a cost (e.g., Hillis, Ernst, Banks, & Landy, 2002). When signals are integrated (or associated), access to the single cue information

may get lost so that a small discrepancy between the two signals cannot be detected anymore. Depending on the task, either the benefit or the cost of integration shows itself. To perform a quantitative analysis of learning, we here used a discrimination task exploiting the cost of sensory integration.

But how does the brain know which sensory signals to integrate? It would make sense to integrate sensory signals only if they come from the same object property or event, and they should have some unique relationship so that when one signal is known, the other one can be inferred. This is a form of “correspondence problem”: Our brain has to know which sensory signals correspond to one another in order for the appropriate signals to be combined. Knowing the correspondence means to know the mapping between the sensory signals. For example, without knowing the mapping between the retinal activation that makes up the visual size estimate and the forces on the finger tip that give rise to the haptic size estimate, it would be impossible to integrate the visual and haptic size signals. Consequently, if the system did not know the mapping between the sensory signals, they could not be integrated and, thus, they had to be treated as independent.

But how can the system know the mapping between the sensory signals? The felt and seen size of an object are two totally different sensory signals: One is derived from photons on the retina, and the other one is derived from sensors detecting the fingers’ position given some force when in contact with the object. However, when simultaneously seeing and feeling an object, there is a natural statistical relationship between these neural signals giving rise to the felt and seen size of an object. This statistical relationship between these totally different sensory signals potentially could have been exploited by the developing brain to form its own concept of the objects’ property, which we now uniquely call “the object’s size.”

There are several studies on within- and cross-modal statistical learning (e.g., Conway & Christiansen, 2005, 2006). However, the question we address here is whether such a new mapping (i.e., a unique correspondence between different signals) can be learned. That is, is it possible to learn to integrate two arbitrary signals, which usually do not correspond in the natural environment and, therefore, are usually not integrated? Phrased differently, we ask whether the integration of sensory signals is predetermined and hardwired in the nervous system or whether it is adaptive.

We decided to test this using visual and haptic perception. We chose the luminance of an object as the visual dimension and its stiffness as the haptic dimension. This choice was made because we believed that there should be no statistical relationship between these two properties in a “natural” environment. By correlating these properties (and hence the visual and haptic signals derived from them) in an artificial environment, we can introduce a new statistical relationship between these two signals. A comparison of discrimination performance before and

after extensive training with such correlated stimuli will reveal whether cue integration can be learned from the statistical co-occurrence of the stimuli in the environment.

Bayesian integration model

Bayesian estimation theory provides a principled approach to handle such questions (e.g., Mamassian, Landy, & Maloney, 2002; Yuille & Bülthoff, 1996). There is a recently published Bayesian integration model that we use and extend here to describe learning to integrate two previously unrelated signals (Bresciani, Dammeier, & Ernst, 2006; Ernst, 2005; Jäkel & Ernst, 2003; Roach, Heron, & McGraw, 2006; Shams, Ma, & Beierholm, 2005). In our experiment, we manipulated the correlation between haptic and visual properties of an object. A perceptual system can potentially learn to exploit this correlation between the stimulus properties. Beliefs or knowledge about the joint stimulus statistics is implemented in the form of priors in the Bayesian framework. That is, the prior distribution describes what stimulus combinations the subject expects to encounter. Learning in this view is then reflected in a change of the prior beliefs about the joint distribution of the stimuli (as represented by the sensory measurements; cf. Adams, Graf, & Ernst, 2004).

In the experiment, subjects were presented with physical stimuli having a visual and haptic property (luminance and stiffness): $s = (s_V, s_H)$. Assuming that the sensory measurement $\hat{s} = (\hat{s}_V, \hat{s}_H)$ derived from the physical property is unbiased but noisy with some Gaussian noise σ_i added independently to each property i ($\hat{s}_i = s_i + \sigma_i$)—for example, due to noise in the neural transmission of the signal—then the joint likelihood distribution $p(\hat{s} | s)$ for vision and touch is a 2D Gaussian with mean s and standard deviations σ_i . These σ_i are given as elements of the diagonal 2×2 variance–covariance matrix Σ :

$$p(\hat{s} | s) = N(\hat{s} | s, \Sigma) \text{ with } \Sigma = \begin{pmatrix} \sigma_V^2 & 0 \\ 0 & \sigma_H^2 \end{pmatrix}. \quad (1)$$

A schematic illustration of a joint likelihood distribution is shown in the top row of [Figure 1](#).

If a subject has no prior knowledge about the joint distribution of the stimuli in the world, the best estimate for the underlying bimodal stimulus will be given directly by choosing the maximum of the joint likelihood distribution (\hat{s} in [Equation 1](#); left column in [Figure 1](#)).

On the other hand, if the subject had some belief about the joint stimulus statistics, that is, which stimuli are probable to occur together, his or her estimate should then be influenced by this belief. For example, if a subject was

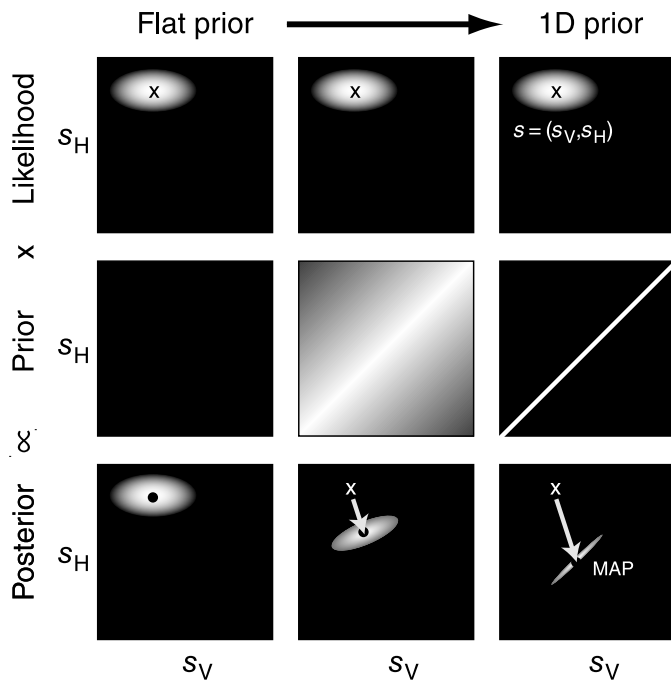


Figure 1. Three schematic examples for combining visual and haptic signals with different priors (columns). Top row: Likelihood distributions with standard deviation σ_V double σ_H ; x denotes physical stimulus. Middle row: Prior distributions; left: flat prior $\sigma_1^2 = \infty$, $\sigma_2^2 = \infty$; middle: $\sigma_1^2 = \infty$, $\infty > \sigma_2^2 > 0$; right: $\sigma_1^2 = \infty$, $\sigma_2^2 = 0$. Bottom row: Posterior distributions, which are the product of the likelihood and prior distributions. The MAP estimate is indicated by \bullet . The arrows indicate the bias in the MAP estimate relative to the physical stimulus (x).

completely certain that there is a perfect relationship between the two properties—for example, light objects are always soft and dark objects are always stiff—it would be enough for the system to measure only one property and to infer the other one. In such a case, the strong belief about the relationship would be represented in a very sharply defined prior $p(s)$, which follows this relationship. Such a sharply defined prior is indicated in the right column of Figure 1. In this example, every pair of sensory signals that does not conform to the known perfect relationship between the two properties would be overruled by the strong prior belief. Hence, an estimate of the underlying stimulus properties should only allow values that are in accordance with the prior distribution.

If the prior belief is not so strong to completely rule out certain estimates (thus, there is some uncertainty in the belief about the relationship), then both the sensory signals and the given prior beliefs should show some influence on the final estimate (middle column in Figure 1). The interplay between the noisy sensory signals and “strength” of the belief about the joint stimulus distribution can be nicely formulated using Bayes’ rule, which states that the posterior is the product of the likelihood distribution—the sensory signals—and the distribution

defining the prior belief, divided by a normalization constant C :

$$p(s|\hat{s}) = \frac{p(\hat{s}|s) \cdot p(s)}{C}. \quad (2)$$

If we assume the prior to be normally distributed $p(s) = N_s(p, \Pi)$ with a mean $p = (0, 0)$ and covariance matrix

$$\Pi = R^T \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} R, \quad (3)$$

in which σ_1^2 and σ_2^2 are the variances of the prior along its principal axis and R is an orthogonal matrix that, in this example, rotates the coordinate system by 45° , then the posterior is also a 2D Gaussian $p(s|\hat{s}) = N_s(s_{\text{MAP}}, \Theta)$ with mean s_{MAP} and covariance matrix Θ .

The maximum of the posterior distribution s_{MAP} is taken to be the estimator for the presented stimulus s . This estimate is called the maximum a posteriori (MAP) estimator. The MAP estimator can be thought of as the optimal way to integrate noisy sensory signals, which are represented by the likelihood function, with extrasensory prior beliefs about the joint stimulus distribution, such as knowledge about the relationship between the physical stimuli. Actually, because such a prior belief is learned from the sensory signals, it can only represent the statistics of the transduced sensory signals and not directly the statistics of the physical stimuli.

The MAP estimator corresponds to a weighted average of the mean of the likelihood and the mean of the prior (given the assumption of all Gaussian distributions):

$$\begin{aligned} s_{\text{MAP}} &= W_\Sigma \hat{s} + W_\Pi p \\ &= \Theta (\Sigma^{-1} \hat{s} + \Pi^{-1} p) \quad \text{with} \\ \Theta &= \frac{1}{(\Sigma^{-1} + \Pi^{-1})}. \end{aligned} \quad (4)$$

The smaller the variance of the prior in a given direction is, the stronger is the influence of the prior in this direction. The smaller the variance of the likelihood function in a certain direction is, the more weight it receives in this direction. This is schematically illustrated in Figure 1: In the left column, we assume that the observer has no knowledge about the mapping between the signals (the uncertainty of knowing the mapping is very big, going to infinity, $\sigma_1^2 \rightarrow \infty$ and $\sigma_2^2 \rightarrow \infty$), so that the prior distribution is completely flat. This is indicated here by the uniformly dark square. The corresponding weights for the prior go to zero in this case. Therefore, the prior should have no influence on the estimates of the physical properties. In other words, the MAP estimate becomes an unbiased maximum likelihood estimate.

In the right column of [Figure 1](#), the prior is very large (close to infinity) along the direction of the correlation between the two properties and close to zero in the perpendicular direction. Zero variance, however, means that the weight for this direction is 1. The estimate, therefore, has to lie on the diagonal that is determined by the perfect correlation between the sensory signals derived from the physical properties. The relative variance of the joint likelihood determines where on the diagonal the MAP will be (Bresciani et al., 2006; Ernst, 2005; see also Roach et al., 2006; Shams et al., 2005). If the system knows with 100% certainty that two properties are perfectly correlated, that is, it knows the mapping between the signals, then only estimates that respect this fact and follow this mapping are allowed. As an example, consider the perception of size. If the system knows that the visually measured size and the haptically measured size of an object are perfectly correlated and it knows the mapping between the measurements, then it can infer that they have to be “identical”; hence, it makes no sense to allow for two separate “percepts of size”—one visual and one haptic. The two measurements of size should be fused to one percept of object size.

In the middle column, there is an intermediate prior between the two extreme conditions—no knowledge about the joint stimulus statistics and perfect knowledge of a linear relationship between measurements. This prior still has a very large (close to infinity) variance along the direction given by the mapping between the properties. However, it has a nonzero variance orthogonal to this direction, indicating that the correlation between the measurements is nonperfect (or it is not known perfectly). As a consequence of such an intermediate prior, which represents some uncertainty in the mapping between the measurements, the MAP estimate is also in between the two extreme cases: It is not completely projected onto the main diagonal; it is only more biased toward the main diagonal as compared to the unbiased estimator in the left panel. Thus, in this case, there is no complete fusion between the signals but only a mutual bias indicating a weaker form of integration between the signals. This weaker form of integration may be called “coupling.” Therefore, Ernst (2005) termed this prior, which results in more or less coupling between the sensory signals, “Coupling Prior.”

Learning to integrate signals

Learning to integrate signals in the Bayesian integration model as presented here is represented as a change of the subject’s belief about the distribution of the stimuli, which is reflected in a modification of the priors. In our experiment, we are manipulating the joint distribution of the stimuli. Before learning, visual and haptic properties

were uncorrelated, and during learning, they have been correlated. Hence, if we observe some effect of learning, it can be attributed to a change in the stimulus statistics.

This can be illustrated with two stimuli, which we assume to be unrelated and, thus, independent—for example, the luminance of an object and its stiffness. Whether an object reflects much light or not most probably does not tell anything about how hard or soft it feels. Therefore, in a “natural” environment, it does not make sense to integrate the sensory signals elicited by these two stimuli. However, if we lived in a world where bright objects always feel hard and dark objects soft, it would suddenly make sense for our sensory system to integrate the visual and the haptic signals. In other words, if the value of one variable was informative about the value of the other (i.e., there is redundancy), it would be useful to integrate these signals. The system can do so given the fact that it knows the relationship between the signals.

Being born in a world where luminance and stiffness are independent and then being put into a situation where these two properties are highly correlated, what is changing? Learning that two properties (or measurements) are highly correlated should change the subject’s belief about the joint distribution of the two sensory signals related to luminance and stiffness. In the Bayesian framework, this is equal to a change in the prior distribution. That is, acquiring more information about the correlation is increasing the certainty in the mapping estimate. As a consequence, the signals should be integrated.

Predictions for discrimination

To test the hypothesis that introducing a correlation between cues should change the prior distribution, we measured subjects’ discrimination performance for objects that can vary in luminance and stiffness before and after extensive training with stimuli for which these two properties were highly correlated. Given the Bayesian model, we can make some predictions of how the discrimination performance should change.

Let us assume that the variance in both cues is equal (we can always rescale the axes such that this holds) and that their noise distributions are independent. Then, the likelihood function is symmetrical with circular cross section in the (s_V, s_H) cue space. If subjects have no prior knowledge about the correlation between the cues, which would imply that the prior is flat, there is no particular direction in this cue space that is different from any other direction. Therefore, the discrimination performance also has to be the same in all directions of this space. This situation is illustrated in the left panel of [Figure 2](#).

On the other hand, if subjects were 100% sure about the mapping between the signals, this is expressed in a prior

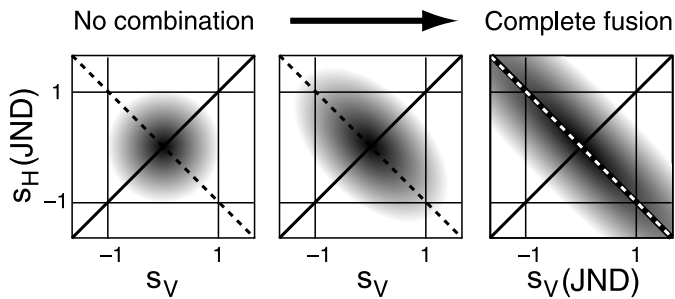


Figure 2. Hypothetical discrimination performance using the MAP estimator. Black corresponds to discrimination performance according to chance level. White corresponds to perfect discriminability. Three examples for discrimination performance of MAP estimates as a result of different priors are shown. Left panel: flat prior corresponding to left row in Figure 1 resulting in equal discrimination performance in all directions; middle panel: intermediate prior resulting in an asymmetric decrease in discrimination performance; right panel: delta function prior resulting in indiscriminability of the fused stimuli (direction of metameric performance).

that is aligned along the direction of correlation (here, the positive diagonal in the cue space) and has a zero variance in all other directions. That is, the prior is a delta function aligned along the positive diagonal. Given such a prior, a stimulus off this diagonal would be fused and is consequently perceived as a stimulus from this diagonal. This can be illustrated with an example: Imagine the Cue 1/Cue 2 stimulus is $+1/-1$ (arbitrary units) and it is fused to be perceived as 0. Then, another stimulus such as, for example, $+2/-2$ may also be fused to be 0. As a consequence of fusion, these two stimuli cannot be discriminated anymore because they both give rise to the same percept. This is the cost of integration. Such stimuli, which are physically different but perceptually equal, are called metamers (Hillis et al., 2002). Given a joint likelihood distribution with equal variance of the two cues, all metamers lie on the negative diagonal in this cue space as is indicated in the right panel of Figure 2 (if the two cues' variance were not the same such as in the example given in Figure 1, the metamer direction would be different and analog to the direction of the arrow in Figure 1; see Bresciani et al., 2006; Ernst, 2005). Thus, given a delta function prior, discrimination thresholds approach infinity in the direction that is defined by the metamers in this space (here, the negative diagonal).

The middle panel of Figure 2 illustrates an intermediate case. Here, the prior is also aligned along the positive diagonal but now has a variance that is different from zero and is less than infinity. Therefore, discrimination performance is impaired in all directions except for the direction defined by the relationship between the signals. Maximal impairment is in the direction of metamers. Discrimination performance does not change in the direction of correlation (positive diagonal) because the

variance of the prior is always approximately infinite in this direction.

That discrimination performance becomes worse when subjects learn to integrate (or associate) signals may seem counterintuitive at first. This is because it is often implied that there should always be a benefit from integration. As can be seen here, however, this intuition is wrong because there is also a cost involved in integrating signals. This cost, as a signature of integration, can be revealed using a discrimination task for which detecting the discrepancy between the signals becomes important.

In summary, if the subjects believe that the properties are uncorrelated, they will effectively carry out the discrimination independently for the two cues (Figure 2, left panel). If they believe that the properties are correlated, they will integrate them (Equation 4), resulting in a cost for discrimination of the stimuli along all directions except for the axis of stimulus correlation (Figure 2, middle and right panels). Learning in this experiment would therefore be manifested in a change in discrimination performance when comparing the performance along the negative and positive diagonals of the cue space. Such a change in relative discrimination performance between these two axes would indicate learning of a Coupling Prior and, thus, learning to integrate arbitrary signals. Thus, we predict to find an interaction between the factors pre-/posttest and congruent/incongruent (negative/positive diagonal) if subjects can learn to integrate arbitrary signals.

Methods

Participants

Twelve trained observers (seven men and five women; 26.1 ± 3.1 years) participated in the experiment for payment. All had normal or corrected-to-normal vision and no history of somatosensory disorders. All were naive to the purpose of the experiment except for the three observers C.R., M.E., and F.J. Subjects were randomly assigned to the different experimental conditions. Participants gave their informed consent before taking part in the experiment.

Setup

To generate the visual and haptic stimuli, we used a mirror setup as depicted in Figure 3. Participants looked onto a mirror and saw a visual scene that is generated on a computer screen. Below the mirror, a subject's index finger was attached to a robot arm with 6 *df* and force feedback along the three translatory directions (PHANToM 1.5, SensAble Technologies, Inc.). Subjects have a convincing

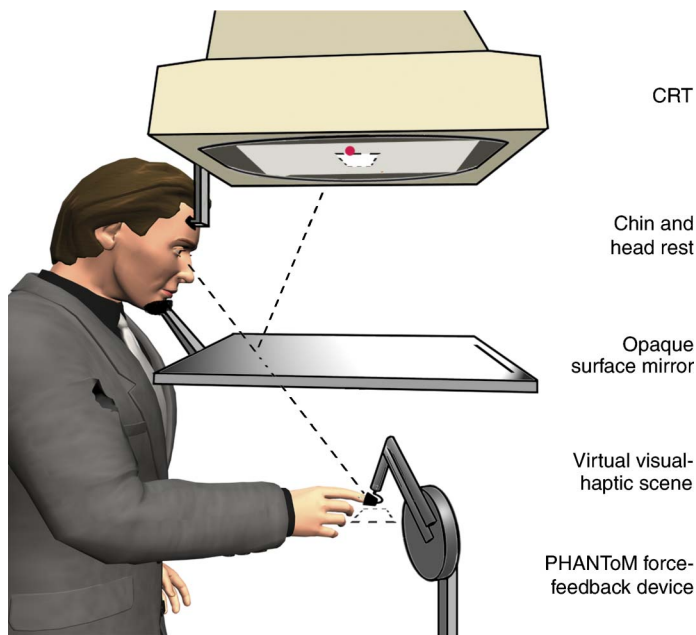


Figure 3. The setup used can display visual scenes on a cathode ray tube (CRT), which are mirrored to be aligned with the haptic scene. Both scenes can be controlled independently. Haptically, the scene can be explored using a PHANTOM device to provide the appropriate force feedback. The subject's head is fixed on a head and chin rest. We used an SGI, Octane 2 to drive the visual and haptic simulation. GHOST was used to generate the haptic scene; OpenGL with GLUT was used for visual rendering.

impression that they are haptically exploring the same scene they are seeing. For details about the setup, see Ernst and Banks (2002).

Stimuli and basic task

Stimuli are flat squares (25×25 mm) viewed at a distance of approximately 50 cm that can have a haptic and a visual property, namely, a certain stiffness and luminance. All other properties are kept constant throughout the experiment.

- The stiffness of the square is modeled using a linear spring model with spring constant k (GHOST, SensAble Technologies, Inc.). The maximum stiffness that can be reliably generated with this device is $k = 0.65$ N/mm. That is, we used a stiffness ranging from 0 to 0.65 N/mm. The range is normalized from 0 to 1; hence, the maximum $k = 0.65$ corresponds to 1.
- For the luminance, we only used the green electron beam (Sony Trinitron F500R). The exponent of the gamma correction was predetermined with a photometer (Minolta). We were able to present 1,024 different shades of green. We normalized the range from 0 to 1; hence, the maximum luminance 58 cd/m² corresponds to 1.

There were three presentation conditions: haptic alone, vision alone, and visual–haptic. To measure discrimination performance in all three presentation conditions, we used a three-interval forced-choice (3-IFC) oddity task. We chose an oddity task because it allows to quantitatively model the discrimination results and because it is not susceptible to criterion biases such as, for example, a 2-IFC same/different task would be. Subjects sequentially saw and/or felt three objects (little squares) from the two-dimensional visual–haptic stimulus space. Each square was presented for 500 ms. Two of the stimuli were identical and one was different in some aspect (luminance, stiffness, or both). Their task was to identify the interval containing the odd stimulus (Figure 4).

The presentation procedure was as follows: A white outline of the first square appears randomly on 1 of 16 possible locations. The subject reaches out for the square with one finger. In the haptic-alone condition, once the subject reaches the square, he or she receives a sensation of stiffness for 500 ms. In the visual–haptic condition, the square lights up with a certain luminance simultaneously with the haptic stimulus for 500 ms. In the visual-alone condition, only the square lights up, but the subject reaches through the square without haptic feedback. There was a 250-ms interstimulus interval before the outline of the next square at a different location indicated the start of the next stimulus presentation. After the presentation of the third stimulus, the subjects made their choice by pressing their finger on one of the three stimulus locations (the locations were indicated by outline squares with the respective number of the interval written on them so that they were easy to identify).

To measure the just-noticeable differences (JNDs) using the oddity task, we adopted a constant stimuli procedure with a fixed standard that was the same for all subjects and conditions (all sessions except training). The standard luminance value and standard stiffness value were chosen to be in the middle of the logarithmic stimulus range that we could present (0.15 N/mm, 13 cd/m²). Each trial consisted of

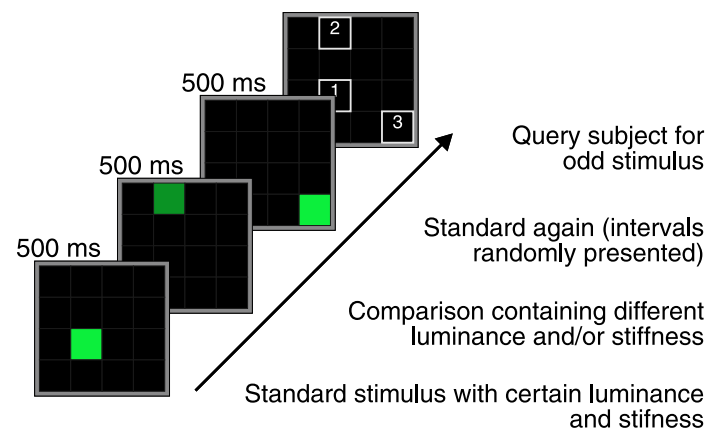


Figure 4. Schematic illustration of the vision-alone oddity task. The procedure during the haptic-alone task or the visual–haptic task was the same but with haptic stimulation either alone or simultaneously with the visual stimulation.

the fixed standard and a comparison stimulus differing in luminance and/or stiffness from the standard. The odd stimulus, which was the stimulus that was only presented once during a trial, could be either the standard or the comparison stimulus chosen randomly with equal probability. To avoid participants learning the standard stimulus, we have included trials (which make up 10% of the total number of trials) where the standard stimulus was not shown; these trials were discarded for the analysis.

Analysis

Plotting error rate against the difference between standard and comparison (in log units), to a very good approximation, the discrimination data have a Gaussian shape. It is not immediately clear what shape of psychometric function to expect from the oddity task if subjects used the MAP estimator that we described in the [Introduction](#) section for making the discrimination. Hence, we have simulated the task with Gaussian noise added to the stimuli and using the triangular rule for the oddity task as decision rule (Versfeld, Dai, & Green, 1996). Using the triangular rule, the estimate that is furthest away (in the Euclidian sense) from the center of the other two is guessed to be the odd stimulus. The simulation revealed that the resulting psychometric function is well described by a Gaussian. Because we used a three-interval oddity task, the chance level for the error rate is at 66%. In none of the conditions tested did subjects' discrimination performance deviate significantly from the Gaussian prediction.

By fitting a Gaussian to the log of the discrimination data using a maximum likelihood fitting procedure, we defined the threshold θ to be 1 *SD* of this Gaussian. Besides the standard deviation, we also had a nuisance parameter λ to account for non-task-related observer lapses (Wichmann & Hill, 2001).

Conditions

The experiment had a two-factor within-subject design. Each subject performed a pre- and posttest with a training phase in between. Thus, one factor was performance before and after training. During training, subjects were exposed to correlated stimuli only. In pre- and posttest, stimuli came either from a correlated or from an anticorrelated distribution (two directions in visual–haptic space; blue diagonal for correlated and red diagonal for anticorrelated in [Figure 6](#)). That is, the second factor was the congruent (blue) or incongruent (red) direction relative to the correlation during training. The dependent variable is the discrimination performance (JND) in the four conditions (pre/post; congruent/incongruent).

The experiment was divided into five sessions conducted on four separate days. Each session lasted between 1.5 and 2.5 hr.

First day: Session 1—Normalization

In a first session, we individually determined a subject's single-cue JNDs in a purely visual and a purely haptic discrimination task. In the purely visual task, we varied the luminance of the squares, but the squares did not give any force feedback. In the purely haptic task, the squares had some stiffness when touched, but visually, there was only the white outline visible. To measure a psychometric function for both tasks, 12 comparison stimuli were chosen logarithmically over the whole stimulus range that we could present. One of the comparison stimuli was randomly chosen each trial. Each stimulus pair of standard and comparison was measured 25 times. Thus, a psychometric function in the unimodal tasks contained 300 decisions per observer. The order of the visual and haptic task was balanced over participants. An example of a maximum likelihood fit to the discrimination data of one subject (D.C.) in the purely haptic task is shown in [Figure 5](#). The standard deviation of the fitted Gaussian is taken as the individual visual and haptic JND.

Second day: Session 2—Pretest

With the knowledge about individual visual and haptic JNDs for each subject, we could normalize the stimulus space individually for each subject in units of JND. We measured discrimination performance for the two-cue stimuli along two directions in this (individually)

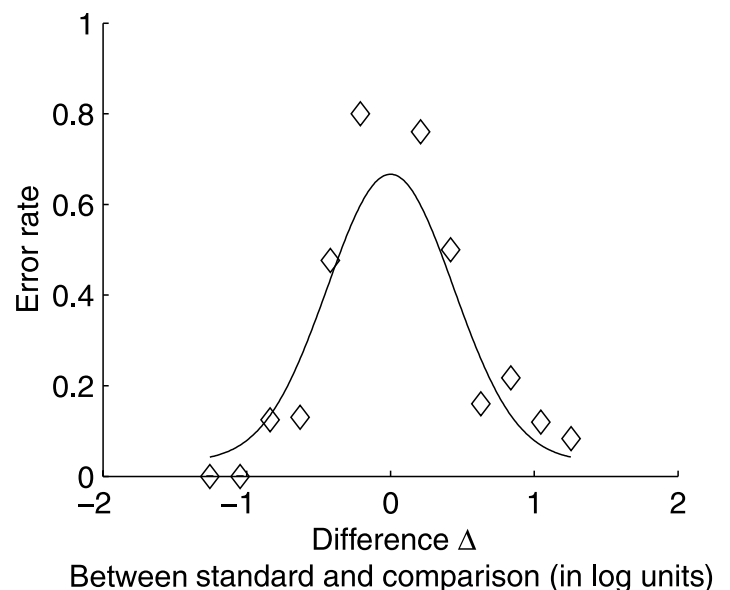


Figure 5. Discrimination data from one subject (D.C.) in the oddity task with only haptic information available. Plotted is the error rate for identifying the odd stimulus versus the difference in stiffness between standard and comparison stimulus in log units with the fixed standard shifted to zero. We measured 25 repetitions per data point.

normalized visual–haptic space. One direction is the positive diagonal, that is, the $(+1;+1)$ axis shown in blue in Figure 6, and the other direction is the negative diagonal, that is, the $(+1;-1)$ axis shown in red.

To measure discrimination performance along these two directions, we again used the oddity task as described above. For both directions, we used the same fixed standard as before, indicated as $(0, 0)$ in this normalized cue space. The comparison stimulus came either from the blue $(+1;+1)$ axis or the red $(+1;-1)$ axis. There were 10 comparison stimuli chosen individually for each subject. These were randomly chosen for each trial to cover a range of ± 2.5 (unimodal) JND units in steps of 0.5. Again, each stimulus pair of standard and comparison was measured 25 times. Thus, each of the two psychometric functions along the positive and negative diagonal contained 250 decisions per observer. Trials from both directions were randomly intermixed. As in the previous case, the odd stimulus could be either the standard or the comparison stimulus.

To get a measure for the discrimination performance, we fit Gaussian psychometric functions to the discrimination data of both directions individually. Thus, we have 1 SD parameter for the threshold in the congruent direction θ_c and one for the threshold in the incongruent direction θ_i . However, we used common lapse rate parameter for the two directions because data for both directions came from the same session.

Figure 7 (upper panel) shows data for one subject (D.C.) split into congruent (blue) and incongruent (red) trials together with a maximum likelihood fit. The best fitting

Gaussians are exactly on top of each other, indicating that discrimination performance in both directions was identical, as would be expected from two unrelated stimulus properties (see also prediction in Figure 6—Panel 2).

Third day: Session 3—Training

During training, we presented stimuli from either only the $(+1;+1)$ axis or the $(+1;-1)$ axis depending on the group that the subject was randomly assigned to. For each group, we will call the direction that the subject was trained on “the congruent direction” and the other one “the incongruent direction.” For each of the two groups, the stimuli were equally distributed along the respective direction spanning the entire possible range. That is, the intensity ranges from close to zero up to approximately the maximum we can present physically. Thus, we choose the widest possible range for the distribution of the stimuli to facilitate learning of the correlation. For each trial, two composite stimuli were chosen randomly from this distribution, and one of them was assigned to be the odd stimulus. During training, subjects received feedback on each trial in the form of a beep that indicated incorrect answers. Each subject performed 500 trials during this training session. It usually took subjects about an hour to complete the training session.

The hypothesis is that subjects learn during training that the variance σ_2^2 along the incongruent axis is reduced compared to stimuli without correlation (before training). At the extreme end, the subject could learn that there is no

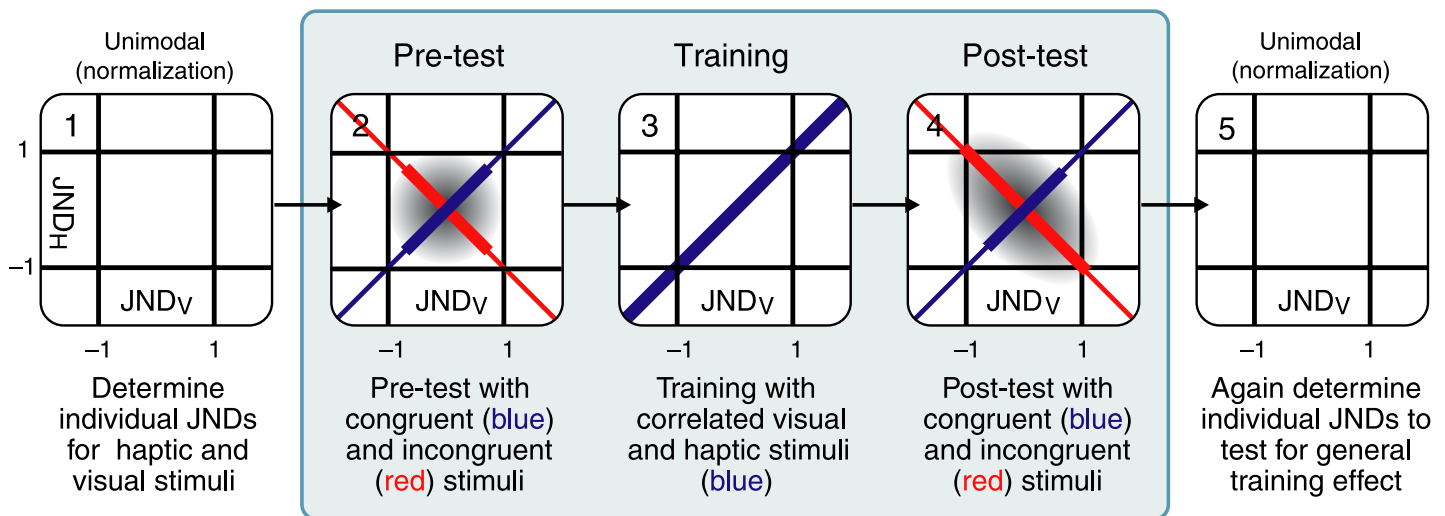


Figure 6. Procedure: (1) Determine JNDs for stiffness and luminance individually using the oddity discrimination task (Day 1). (2) Pretest at Day 2, determining bimodal discrimination performance along the congruent and incongruent directions. (3) Training with correlated bimodal stimuli from the congruent direction (Day 3). (4) Posttest same as pretest directly following training at Day 3. (5) Again, determine individual JNDs for stiffness and luminance at Day 4 (same as Step 1). The light blue box indicates the bimodal conditions from the main experiment.

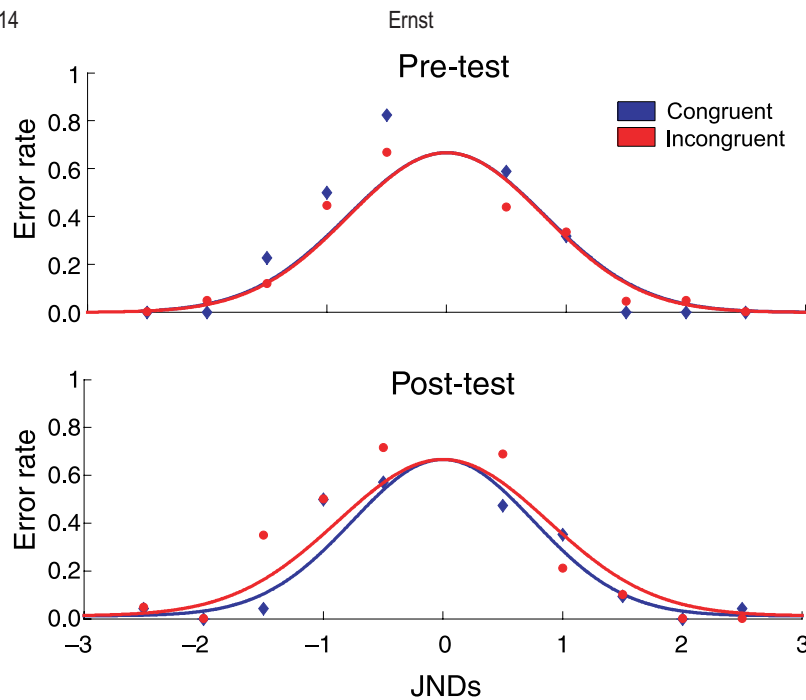


Figure 7. Discrimination performance for subject D.C. for congruent and incongruent trials before and after training (upper panel, pretest; lower panel, posttest). Error rate is plotted against the difference (Δ) between comparison and fixed standard stimulus (given in JND units). Discrimination data plus Gaussian fit for the congruent trials are depicted in blue. Data and fit from incongruent trials are depicted in red (chance level performance at 66%).

variance along the incongruent axis, which would mean that the subject believes that the two signals are completely correlated. After which, subjects had a brief break of a couple of minutes before they immediately continued with the posttest on the same day.

Session 4—Posttest

The procedure of the posttest was identical to that of the pretest with one exception. As it is expected that, during the posttest, where also half the stimuli came from the incongruent distribution, subjects will slowly unlearn what they supposedly had learned during training, we included a number of training trials (one third of all trials), which all came from the congruent direction (here, during posttest, there was no feedback given for the intermixed training trials). Thus, there were 500 regular trials (250 congruent and 250 incongruent) plus 250 training trials, which all came from the congruent direction. Figure 7 shows the psychometric functions for subject D.C. determined as before. Here, a difference between the congruent and incongruent directions is visible, indicating some learning in this one subject.

Fourth day: Session 5—Control

Because subjects now had a lot of experience with the task and the stimuli, it may be that there is perceptual learning; thus, subjects generally get better in performing this task. The last day was made identical to the first day

(normalization) to control for this general learning. We again measured performance in the task with only visual or only haptic information available.

Results

First, we checked whether there was a significant change of the unimodal thresholds over the time course of the experiment (between Session 1 from Day 1 and Session 5 from Day 4). It could be that subjects become much better in discriminating stimuli simply because they have done more than 2,500 trials (perceptual learning). We can compare the discrimination performance in the purely visual task and the purely haptic task on the first day with the performance on the last day. A repeated measures ANOVA shows no significant effect for the purely visual task or the purely haptic task, $F(1, 8) = 0.34$, $p = .56$ and $F(1, 8) = 3.13$, $p = .12$, respectively. This indicates that subjects did not generally get better at discriminating during the course of the experiment. Given this baseline performance, we can now turn to the main data of the experiment.

The results for one typical subject (D.C.) were already shown in Figure 7. For all subjects, we performed the same procedure of fitting a Gaussian to the data in both the congruent and incongruent directions in the pre- and

posttest. Thus, we had thresholds for all four conditions with the two factors: pretest/posttest and congruent/incongruent.

There was no significant difference between the congruent and incongruent direction during pretest in 9 of the 12 subjects. This result was expected from subjects where the two cues are independent and where there is no mapping between these signals. The results of these nine subjects are shown in Figure 8. The top panel shows the mean thresholds for these nine subjects for the congruent and incongruent directions in the pre- and posttest.

An ANOVA (two factors, within subjects) on the data of these nine subjects was conducted and revealed that there was no significant main effect, neither for pretest versus posttest, $F(1, 8) = 0.705$, $p = .426$, nor for congruent versus incongruent, $F(1, 8) = 4.128$, $p = .077$. However, it is important to note that we found a significant interaction between the two factors, pre/post vs. congruent/incongruent: $F(1, 8) = 14.58$, $p < .005$, which indicates that the thresholds for the congruent and incongruent directions, which were the same in the pretest, are now different in the posttest. This shows that these subjects learned to use the newly introduced redundancy between the luminance of an object in combination with its stiffness. That is, subjects learned to integrate arbitrary signals.

In the lower panel of Figure 8, we illustrate the interaction by plotting how the difference between the congruent and incongruent directions change from pre- to posttest. With the exception of one subject, all of the nine subjects show a change in the predicted direction, and the one who did not show the effect has only a minor change in the opposite direction. Thus, this illustrates the significance of the interaction.

Interestingly, 3 of the 12 subjects who were tested (C.R., 0.165; S.L.S., 0.169; V.E., 0.177) already showed a significant difference between the congruent and incongruent directions during pretest before the training. This indicates that they already had a predefined axis for discrimination. (By chance, all three had a higher threshold for the congruent direction compared with the incongruent direction—because the assignment to the group according to which direction a subject was trained was random, this has to be chance.) According to the Bayesian integration model, this could be due to correlated noise in the two channels, which is unlikely because luminance and stiffness are sensed by two entirely separate sensory systems (vision and touch) in these subjects (see the Discussion section for details). More likely, it is that these three subjects lived in a world where there was a slight correlation between brighter objects feeling stiffer or vice versa. If some correlation like this exists somewhere in the environment, these three subjects may have picked that up during their lifetime before the experiment. This is interesting but not so important for our experiment. What

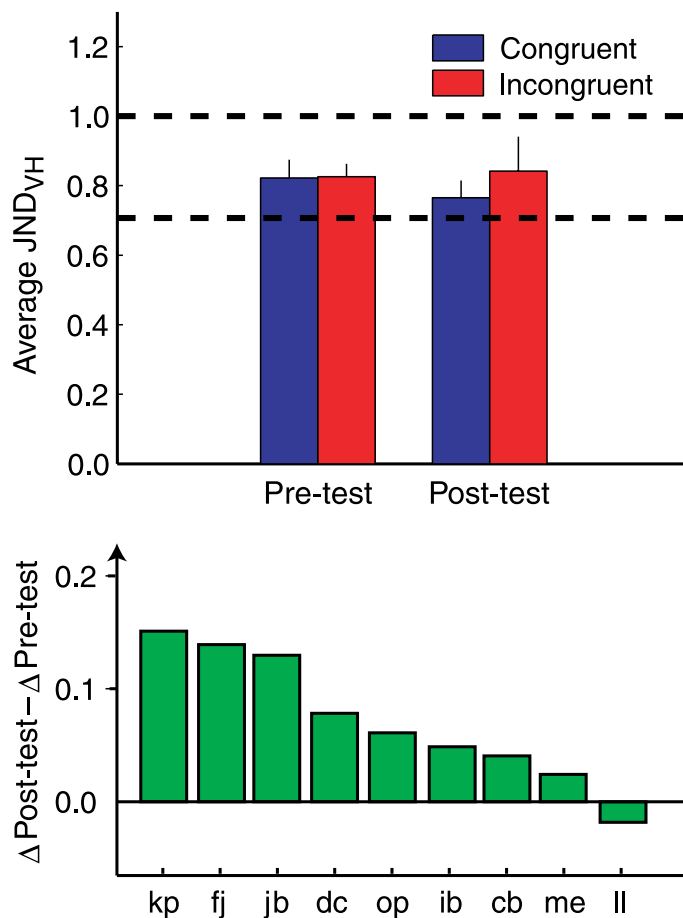


Figure 8. Upper panel: Mean data of discrimination thresholds across the nine subjects in all four conditions plus standard deviation (pre/post, congruent/incongruent) are shown. The upper dashed line at 1.0 represents 1 unimodal JND. The lower dashed line at $1/\sqrt{2}$ (circle with radius 1 JND measured along diagonals) represents the best discrimination performance that can be theoretically achieved along the diagonals. Lower panel: This shows the difference between pre- and posttest of the discrepancy between discrimination thresholds derived from the congruent and incongruent directions.

is important is that these three subjects also produced a similar learning effect as the other subjects ($\Delta \text{pretest} - \Delta \text{posttest}$: C.R. = 0.165, S.L.S. = 0.169, and V.E. = 0.177). The difference between congruent and incongruent trials also changed in the predicted way after learning: The difference became less or disappeared. That is, the learning effect is in the same direction as the eight of the nine subjects shown in the lower panel of Figure 8. Thus, 11 of the 12 subjects showed the predicted effect that one would expect if subjects learned to use the correlation (mapping) introduced during the training phase. Thus, in conclusion, this experiment showed that subjects learned a new Coupling Prior, which enabled them to integrate arbitrary signals from vision and touch.

At the end of the experiment, we informally queried subjects and there was no naive subject who reported to have noticed the correlation during training.

Discussion

Here, we tested whether subjects can learn to integrate arbitrary signals from vision and touch—namely, the luminance and stiffness of an object. We measured discrimination performance for these two signals presented simultaneously and explored whether there is a change in discrimination performance before and after extensive exposure to a world in which these two signals are highly correlated. We assumed that the thresholds should be symmetric before training; hence, there is “no fusion” between these two signals. We then predicted that subjects’ thresholds should become asymmetric after training if they were sensitive to the correlations in the stimuli during training. That is, the signals should become “somewhat fused,” which we here call integration. In case subjects’ priors were fully adapted, one could say that the amount of correlation in the sensory measurements derived from the physical stimuli determines the “degree of fusion.” Thus, the prior, which we termed Coupling Prior, represents the mapping uncertainty (Bresciani et al., 2006; Ernst, 2005).

Our main finding is that subjects actually showed the predicted learning effect. For most of the subjects (9 of 12), there was indeed no difference between the discrimination thresholds in the congruent and incongruent direction before training, indicating independence of the signals. All except one subject ($n = 11$) were sensitive to the training and showed the predicted learning effect. That is, after training, there is a larger difference in thresholds in the incongruent direction than in the congruent direction. This suggests that subjects indeed learned to integrate the two arbitrarily chosen signals—luminance and stiffness. The asymmetry between congruent and incongruent thresholds cannot be explained by improvement of performance due to more practice because this would have affected the congruent and incongruent direction equally. Furthermore, we controlled for such unspecific learning by measuring the unimodal discrimination performance before and after the pre- and posttest and found no significant difference.

As long as we assume that the noise distributions of the luminance and stiffness measurements are independent, there is no way that changing the likelihood distribution (either by introducing a bias to its mean or by changing its variance σ_V^2 or σ_H^2) would have produced an asymmetry in the discrimination performance between the congruent and the incongruent direction. This independence assumption of the noise distributions of the two sensory measurements seems safe because the measurements are derived from two separate sensory modalities.

Furthermore, there is no reason to believe that introducing a correlation between the signals during training would affect this independence assumption of the noise distributions of the signals. Thus, the asymmetry in the learning effect between congruent and incongruent direction can be best explained by a change in the variance of the Coupling Prior and not by a change in the likelihood distribution.

Recalibration is another form of learning involving a conjunction of two sensory signals, which can be modeled using a similar model as the one described here (Burge, Ernst, & Banks, 2007). Recalibration occurs when the system is exposed for some time to a constant conflict between the two sensory signals. The classical example for recalibration is visuomotor adaptation, which has been extensively studied since the first prism experiments conducted by von Helmholtz (1867). Similar recalibration effects have been found within the sensory modalities (e.g., Adams, Banks, & van Ee, 2001). In essence, recalibration affects the mapping between the signals and not the certainty of the mapping or the variance of the sensory measurements (likelihood). Thus, recalibration cannot explain the asymmetry effects in discrimination performance found here.

In the pretest, all subjects showed a (bimodal) threshold below 1 JND (unimodal). If a subject only used the visual or only the haptic signal for discrimination, the threshold should lie at exactly 1 JND. Thus, we can conclude that subjects used both signals for the bimodal discrimination task. However, they did not use the available information in the best possible way. If they used all the information available, then their threshold should lie at $1/\sqrt{2}$. This is expected when there are two signals instead of only one signal available for the discrimination. To be more precise, if the noise in the sensory measurements is radially symmetric everywhere in the normalized visual–haptic space (i.e., the joint likelihood distribution would have a circular cross section), then discrimination performance should be independent of the direction in the cue space in case the sources of information are not integrated (i.e., the prior is flat). Thus, discrimination thresholds should all lie on a circle in this space (with radius 1 unimodal JND). Because the visual–haptic thresholds are measured along the diagonal direction, then $1/\sqrt{2}$ of the visual and haptic JNDs are needed before hitting the threshold. Because the visual–haptic thresholds in the pretest are slightly above this theoretical optimum, it seems that subjects do not optimally use both sources of information for performing in the bimodal discrimination task during pretest.

Because subjects’ performance was not optimal in the pretest, there is still the possibility for subjects to improve the discrimination performance in the congruent direction during posttest. This was not predicted by the Bayesian integration model. However, subjects apparently learned to better use the two sources of information simultaneously during the training task in which we provided feedback. After training in the posttest, subjects

are very close to optimal in the congruent direction. In the incongruent direction, however, there is a cost for discrimination, which is the result of integrating the sensory signals. This was the prediction of the Bayesian model when a nonignorant Coupling Prior is used together with the sensory information for making the discriminations.

The observed effect is small. However, we did not expect the effect to be large, considering only 1 hr of training compared to a whole lifetime of experience with objects, for which luminance and stiffness are not correlated. If the system would adapt more quickly, serious problems could arise. For example, if, by accident, stimuli get fused, this may result in difficulty in discriminating one stimulus from another (this is the cost of fusion and the reason why the discrimination ellipse becomes wider along the negative diagonal).

Along the same line, one interesting fact is that the Bayesian model actually predicts that subjects' discrimination performance should mostly get worse in this particular task, when they learn to use the correlation between the stimuli. Thus, there seems to be only a cost but no benefit when integrating the signals. This can be best seen for the case of complete fusion. There, it is not possible at all to discriminate along the negative diagonal axis. Hence, it seems that for this task, the cost exceeds the benefits when subjects learn to use the correlation. However, whether or not there is a benefit of integration depends on the particular task. For a different task such as magnitude estimation, for example, combining signals has more benefits, as mentioned in the [Introduction](#) section. This makes sense; if the system is certain about the mapping, then it does not need to have access to both signals because they are redundant. Thus, in return, the system can afford to lose the discriminability between these signals. This is what we see in our task. In return, it is able to obtain a more precise estimate of the environmental property that both signals provide information about. Hence, the benefit is that the precision of the estimation increases, which should be a general goal of the perceptual system. Costs and benefits are balanced based on the certainty that the two signals carry redundant information, which is represented in the Coupling Prior.

The Coupling Prior, learned from the sensory signals mediating the statistics of the environment, represents the mapping uncertainty between the sensory measurements. If the mapping between signals varies often and unpredictably (e.g., when the system has to recalibrate during visuomotor adaptation), this should be reflected in a larger variance of the prior distribution. If the mapping is relatively constant, the mapping uncertainty will be less and, thus, the prior distribution has a smaller variance. Hence, this framework can explain why some signals are fused whereas others are not. Objects that look big also feel big, and objects that feel small also look small. We all grew up with this form of statistical relationship. However, in some properties, mapping is more constant than in

others. For example, when there is a conflict between disparity and perspective cues within the visual system, the adaptation rate is very slow (Adams et al., 2001), indicating a relatively fixed mapping. For visual–haptic conflicts, adaptation is much faster, indicating that the mapping is less fixed. It would be suboptimal for our perceptual system not to use the information about mapping certainty.

Robust estimation refers to the effect that integration should break once there is a large (spatial, temporal, or other) conflict between sources of information. This is because it would not make sense to integrate discrepant signals, which, more probably, have come from two different objects, rather than reflecting signals from the same object. This concern was already discussed in Landy et al. (1995) and was also recently addressed by Knill (2003, [in press](#)) and Roach et al. (2006). Landy et al. treated discrepant signals as statistical outliers, whereas Knill proposed a mixture model of different Gaussians to deal with the problem of robust estimation. Roach et al. proposed a similar idea as Knill but in the context of a Coupling Prior. Both approaches may work. However, when implementing robust estimation as statistical outliers, more than two redundant signals are needed to identify the outlier. Often, there are not more than two corresponding signals available at a given time. The use of a mixture of Gaussians to implement robust estimation was studied using stereo and texture perception of slant. However, the specific shape of the mixture of Gaussian distribution needed for the approach to work seemed very specific to these two cues. Thus, the question of how generally applicable these approaches will be remains.

In the framework presented here, robust estimation is incorporated quite naturally. It seems safe to assume that the more discrepant the signals are (in time, in space, or in any other dimension), the less is the correlation between these signals. Thus, the Coupling Prior conditioned on the discrepancy between the signals has more variance the more discrepant the signals are. If they become really discrepant, the correlation disappears and the signals are treated as independent. If the signals are congruent in all dimensions, it is only the uncertainty in the mapping that determines the variance of the prior and, hence, the “degree of fusion.” An example for robust estimation including a Coupling Prior, which shows that as uncertainty increases, the larger the temporal discrepancy between the signals becomes, is illustrated in [Figure 9](#). The idea presented here is not in contrast with the mixture model idea proposed by Knill (2003, [in press](#)) and Roach et al. (2006). Both models could easily be combined into a more complete framework for robustness in sensory combination.

Haijiang, Saunders, Stone, and Backus (2006) recently studied “cue recruitment.” In their study, participants learned to use a secondary, naturally uninformative cue for disambiguation to interpret an ambiguous figure (the Necker Cube). In a learning phase, a statistical correlation

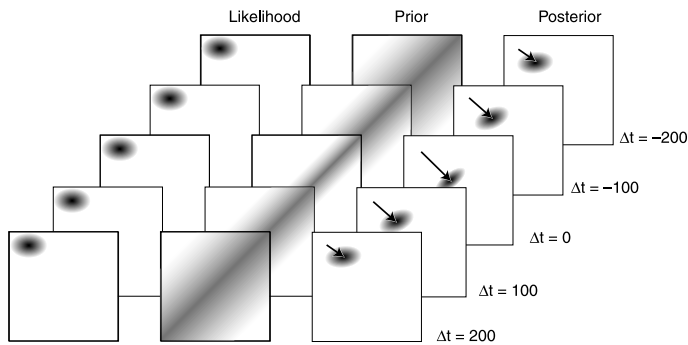


Figure 9. Schematic illustration demonstrating robust estimation. The variance of the Coupling Prior increases with increasing temporal disparity between two signals. Thus, the weight of the prior deteriorates with temporal asynchrony and, hence, the effect of integration disappears. With large temporal asynchronies, the two signals can then be perceived independently of one another.

between the new cue and one or the other interpretation of the ambiguous figure was introduced. The question was whether, after learning, the normally uninformative cue could be used for disambiguation of the ambiguous stimulus. This is a form of association learning that is similar to learning to integrate two sensory signals. However, in the present study, we did not investigate the combination of signals for disambiguation; rather, we examined whether knowledge about the correlation between signals can be used in the combination of the signals.

We will conclude with a citation from Bishop George Berkeley who stated in 1732 in his famous work *Essay Towards a New Theory of Vision*:

“Sitting in my Study I hear a Coach drive along the street; I look through the Casement and see it; I walk out and enter into it; thus, common Speech would incline one to think, I heard, saw, and touch’d the same thing, to wit, the Coach. It is nevertheless certain, the Ideas intromitted by each Sense are widely different, and distinct from each other; but having been observed constantly to go together, they are spoken of as one and the same thing.”

Thus, he already said what we now have experimental evidence of: It is the signals derived from the different sensory modalities that are widely different and distinct from each other, but when there is a statistical correlation between these signals, they are integrated into one thing.

Acknowledgments

I thank Frank Jäkel for help with the conduction of the experiment and the development of the theoretical framework. This work was supported by the IST Program of the

European Commission through the projects “Touch-HapSys” (IST-2001-38040) and “ImmerSense” (IST-2006-027141), by the Human Frontiers Science Program through Grant RGP0003/2006-C, and by the Max Planck Society. An earlier version of the manuscript has been presented at the EuroHaptics 03 conference in Dublin, Ireland, by Jäkel and Ernst (2003).

Commercial relationships: none.

Corresponding author: Marc O. Ernst.

Email: marc.ernst@tuebingen.mpg.de.

Address: Spemannstr. 41, 72076 Tübingen, Germany.

References

- Adams, W. J., Banks, M. S., & van Ee, R. (2001). Adaptation to three-dimensional distortions in human vision. *Nature Neuroscience*, *4*, 1063–1064. [PubMed] [Article]
- Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the ‘light-from-above’ prior. *Nature Neuroscience*, *7*, 1057–1058. [PubMed] [Article]
- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262. [PubMed] [Article]
- 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*(5):2, 554–564, <http://journalofvision.org/6/5/2/>, doi:10.1167/6/5/2. [PubMed] [Article]
- Bülthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America A, Optics and image science*, *5*, 1749–1758. [PubMed]
- Burge, J., Ernst, M. O., & Banks, M. S. (2007). *The statistical determinants of adaptation rate in human reaching*. Submitted for publication.
- Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Boston: Kluwer Academic Publishers.
- Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24–39. [PubMed]
- Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*, *17*, 905–912. [PubMed]
- Ernst, M. O. (2005). A Bayesian view on multimodal cue integration (chapter 6). In G. Knoblich, M. Grosjean, I. Thornton, & M. Shiffrar (Eds.), *Human body*

- perception from the inside out* (pp. 105–131). New York, USA: Oxford University Press.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433. [PubMed]
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162–169. [PubMed]
- 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: Elsevier Science Publishers.
- Haijiang, Q., Saunders, J. A., Stone, R. W., & Backus, B. T. (2006). Demonstration of cue recruitment: Change in visual appearance by means of Pavlovian conditioning. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 483–486. [PubMed] [Article]
- Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research*, *179*, 595–606. [PubMed]
- 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. [PubMed]
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*(12):1, 967–992, <http://journalofvision.org/4/12/1/>, doi:10.1167/4.12.1. [PubMed] [Article]
- Howard, I. P., & Rogers, B. J. (2002). Interactions between depth cues. In I. P. Howard & B. J. Rogers (Eds.), *Seeing in depth* (vol. 2, pp. 469–493). Toronto: I. Porteous.
- Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*, 345–350. [PubMed]
- Jäkel, F., & Ernst, M. O. (2003). Learning to combine arbitrary signals from vision and touch. In I. Oakley, S. O. Modhrai, & F. Newell (Eds.), *Eurohaptics 2003 Conference Proceedings* (pp. 276–290). Dublin 2, Ireland: Trinity College Dublin and Media Lab Europe, Trinity College.
- Knill, D. C. (2003). Mixture models and the probabilistic structure of depth cues. *Vision Research*, *43*, 831–854. [PubMed]
- Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, *7*(7):5 1–24, <http://journalofvision.org/7/7/5/>, doi:10.1167/7.7.5. [PubMed] [Article]
- 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. [PubMed]
- Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture defined edges. *Journal of the Optical Society of America A, Optics, image science, and vision*, *18*, 2307–2320. [PubMed]
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*, 389–412. [PubMed]
- Mamassian, P., Landy, M., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain* (pp. 13–36). Cambridge, Massachusetts, USA: MIT Press.
- Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: A strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the The Royal Society B: Biological Sciences*, *273*, 2159–2168. [PubMed] [Article]
- Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, *16*, 1923–1927. [PubMed]
- Versfeld, N. J., Dai, H., & Green, D. M. (1996). Optimum decision rules for the oddity task. *Perception & Psychophysics*, *58*, 10–21. [PubMed]
- von Helmholtz, H. (1867). *Handbuch der physiologischen Optik*. Leipzig, Germany: Voss.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*, 1293–1313. [PubMed]
- Yuille, A. L., & Bühlhoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123–161). New York: Cambridge University Press.