A Bayesian view on multimodal cue integration

Marc O. Ernst

Max-Planck-Institut für biologische Kybernetik

Spemannstr. 38, 72076 Tübingen, Germany

Tel: +49-7071-601 644 -- Fax: +49-7071-601 616

marc.ernst@tuebingen.mpg.de

in

Human Body Perception From The Inside Out

Knoblich, G., Thornton, I. M., Grosjean, M., & Shiffrar, M. (Eds.) New York, NY: Oxford University Press. Chapter 6, 105-131 (2006)

1 - Introduction

We perceive our own body and the world surrounding us via multiple sources of sensory information derived from several modalities, including vision, touch and audition. To enable interactions with the environment this information has to converge into a coherent and unambiguous multimodal percept of the body and the world. But how does the brain come up with such a unique percept? In this chapter I review a model that in the statistical sense describes an optimal integration mechanism. The benefit of integrating sensory information comes from a reduction in variance of the final perceptual estimate. Furthermore, I point out how this integration scheme can be incorporated in a larger framework using Bayesian decision theory (BDT).

To illustrate the problem of sensory integration, imagine driving a nail into wood using a hammer. The position of the nail in space can be seen, but may also be derived from an estimate of the arm posture, while holding the nail in one hand. That is, vision and the estimate of body posture both provide information about the nails position in space. Slight discrepancies in the representation of information between the estimates naturally arise due to the fact that the process of sensory estimation is inherently noisy. This results in an interesting situation: the observer either has to decide which information to trust in a given situation (vision or the body sense) or it has to find a way to best combine the discrepant information and come up with an optimal decision (or action).

However, having more than one (redundant) estimate available can be an advantage: the accuracy with which an environmental property can be judged increases with the number of individual perceptual estimates available.

In the hammer example, the position in space can be estimated more reliably using both estimates (vision and the sense of the body's posture from somatosensory information) instead of only one. That is, to accurately hit the nail with the hammer it is best to integrate the position information from the two estimates into one common representation (for a more detailed discussion on the integration of sensory information into a body image see Maravita in this volume). One could speculate that this may be one reason it is better for you to hold the nail yourself, instead of having someone else hold it for you while hitting it with the hammer.

Not all information derived from different sensory modalities is redundant. In the majority of cases information derived from the different modalities will be complementary in nature, such as when feeling an object's weight while seeing its colour. Naturally, different combination rules have to be applied for combining such complementary information into a stable percept (for a recent review see Ernst & Bülthoff, 2004). Here, I concentrate on the integration mechanisms for redundant sensory information, such as the spatial position of the hand or the size of an object that can be seen and felt simultaneously.

2 - The probabilistic nature of sensory estimation

The problem of sensory combination can be understood using signal detection theory (Green & Swets, 1988). Perception is a probabilistic process. If one estimates an environmental property, such as an object's size, the estimate will have some variance associated with it. As a result, if the same environmental property is estimated consecutively 100 times, all 100 perceptual estimates may vary slightly. Figure 1 shows schematically the probability density distribution for estimating an object's size *s*. In the simplest case this probability density distribution has a Gaussian shape and is unbiased. This distribution is then defined by its mean $\bar{S}$, which for an unbiased estimator corresponds to the objects size *s*, and its standard deviation σ:

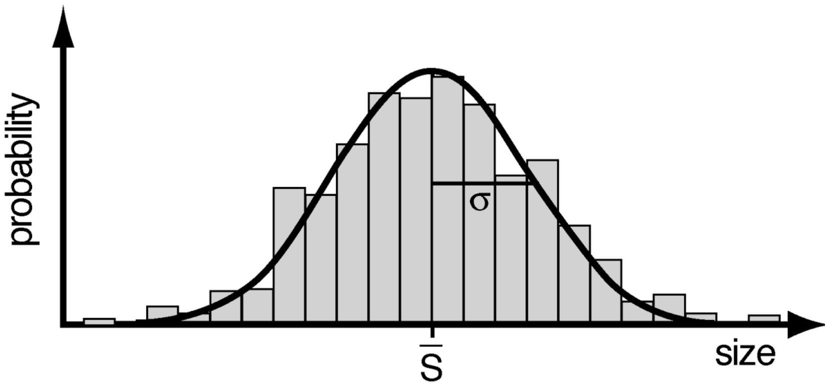
$$\hat{S} = N(\bar{S}, \sigma). \tag{1}$$

If the reliability *r* is defined as the inverse of the variance σ^2

$$r = 1/\sigma^2, \tag{2}$$

then the larger the variance of the associated distribution the less reliable is the associated perceptual estimate.

A)



B)

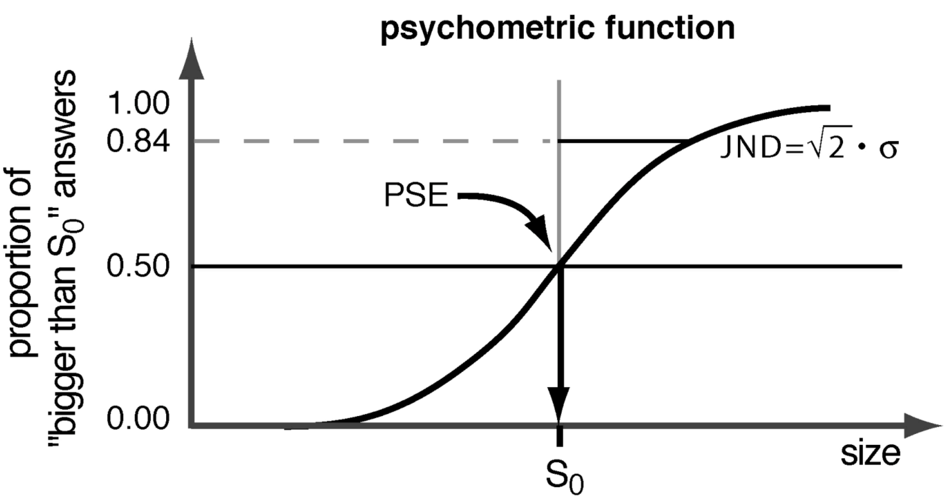


Figure 1: A) Schematic illustration of the probability density function for the estimation of an object's size *s*. The histogram indicates the distribution of answers derived from the size estimation process. The fitted curve has a Gaussian shape (with standard deviation σ and mean $\overline{S}$) and indicates the probability density function. B) Schematic drawing of a psychometric function derived using a 2-interval forced-choice task given the probability density function for estimating the object's size from A. The just noticeable difference (*JND*) derived at the 0.84 point corresponds to $JND = \sqrt{2} \cdot \sigma$. *PSE* is the point of subjective equality.

For experimentally estimating the variance of a sensory signal classical psychophysical discrimination paradigms, such as a 2-interval forced-choice (2-IFC) task, can be used. Subjects performing this task must compare, for example, the sizes of two objects – Standard S_0 with Comparison S – presented sequentially. If the difference in size between the two intervals ($S-S_0$) is large, subjects will have no problem discriminating them, and consequently they will make only few errors. With decreasing size differences however the error rate will rise. If the probability density functions for S and S_0 are Gaussian with identical variance σ^2, the resulting psychometric function is a cumulative Gaussian (see Fig. 1). The "Just Noticeable Difference" (*JND*) defined at the 84% level (the difference between the 50% and the 84% points) provides an estimate

$$JND = \sqrt{2}\sigma \tag{3}$$

for the variability of the underlying Gaussian distribution.

3 - Combining redundant signals

"Redundant signals" may to some degree sound like a waste of information. But actually this is not necessarily so. There are two major advantages in having redundant information available: the first is that the system is more robust, because when one estimate is not available at a given time (or its information is degraded) the other estimate can substitute for it. The second advantage is that the final estimate becomes potentially more reliable compared with the reliability of the individual estimates feeding into the combined percept.

What is the statistically optimal strategy for combining redundant sensory information? Figure 2 shows the probability density distributions (the likelihood functions) for two independent estimates each of which is derived from a stimulus in a different modality. In the example discussed here it is a size estimate that is derived from a visual and a haptic size stimulus ($\hat{S}_V$ and $\hat{S}_H$). According to the "Maximum-Likelihood-Estimation" (MLE) scheme the integrated estimate $\hat{S}_{VH}$ is a weighted average across the individual sensory signals with weights w_i that sum up to unity (the index i refers to the individual modalities $i=1\ldots j\ldots N$) (Cochran, 1937):

$$\hat{S} = \sum_i w_i \hat{S}_i \qquad \text{where} \qquad \sum_i w_i = 1 . \tag{4}$$

Optimally, weights are chosen to be proportional to the reliability of a given signal. That is, if the visual modality provides the more reliable information in a given situation, this signal is given higher weight.

$$w_j = \frac{r_j}{\sum_i r_i} . \tag{5}$$

In the example shown in Fig. 2 the variance associated with the visual size estimate is four times less than the variance associated with the haptic size estimate. That is, the visual information is four times more reliable. Therefore, the combined estimate (the weighted sum) is closer to the visual than the haptic estimate (in the example here the visual weight is 0.8 according to Eq. 5). Under other circumstances where the haptic modality provides the more reliable estimate the situation is reversed.

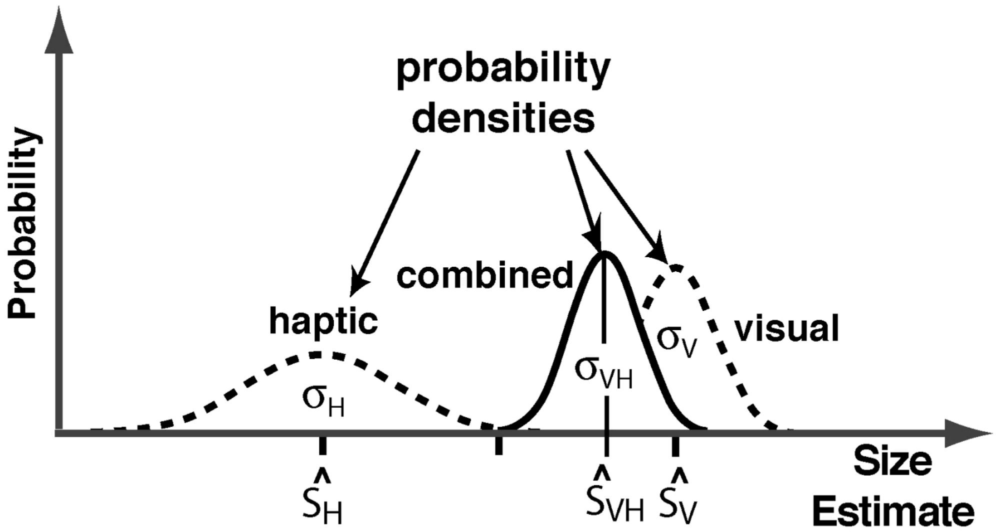


Figure 2: Schematic drawing of the likelihood functions of the individual visual and haptic size estimates and of the combined visual-haptic size estimate, which is a weighted average according to the MLE integration rule. The variance associated with the visual-haptic distribution is less than either of the two individual estimates.

The variance of the combined estimate will be less then that of either of the individual estimates feeding into the combination process. That is, the reliability improves when information is combined. According to the MLE principle the reliability of the combined estimate is the sum of the reliabilities of the individual estimates:

$$r = \sum_i r_i \ . \tag{6}$$

One can show that the MLE integration scheme is statistically optimal in that it provides the most reliable unbiased sensory estimate, given that the individual estimates are Gaussian distributed and that these noise distributions are independent. However, even if the noise distributions of the individual estimates show a correlation one can still benefit from the combination of sensory information and the combined estimate will be more reliable than each individual estimate alone (Oruç, Maloney, & Landy, 2003).

In a recent study we showed that humans integrate visual and haptic information in such a statistically optimal fashion (Ernst & Banks, 2002). Others have demonstrated that this finding of optimality holds not only for the integration across vision and touch, but also for the integration of information across and within other sensory modalities, such as vision and audition (Alais & Burr, 2003; Knill & Saunders, 2003; Hillis, Watt, Landy, & Banks, 2004). Further, the MLE scheme holds also for the integration of sources of sensory information that include the body sense. This was recently shown by van Beers, Sittig, and van der Gon (1998, 1999). They investigated how proprioceptive information about the position of the hand in space integrates with visual information and found the MLE model qualitatively confirmed. That is, perceptual information about one's own body seems to be no different from information derived from the other sensory modalities as far as the integration mechanism is concerned. Thus, maximum-likelihood-estimation is an effective and widely used strategy utilized by the perceptual system.

In the Ernst and Banks (2002) study, subjects had to discriminate the sizes of two objects presented sequentially in a 2-IFC task. The objects could either only be seen, only be felt, or both, seen and felt, simultaneously. The visual stimulus was a random-dot stereogram portraying a bar of given size. The haptic stimulus was generated using two haptic force-feedback devices

(Phantom™ from SensAble Inc.) (see Fig. 3 for details). To vary the reliability of the visual stimulus we added noise to the depth of the dots that formed the random-dot pattern (0%, 67%, 133% 200% depth noise relative to the depth the bar was raised from the background plane).

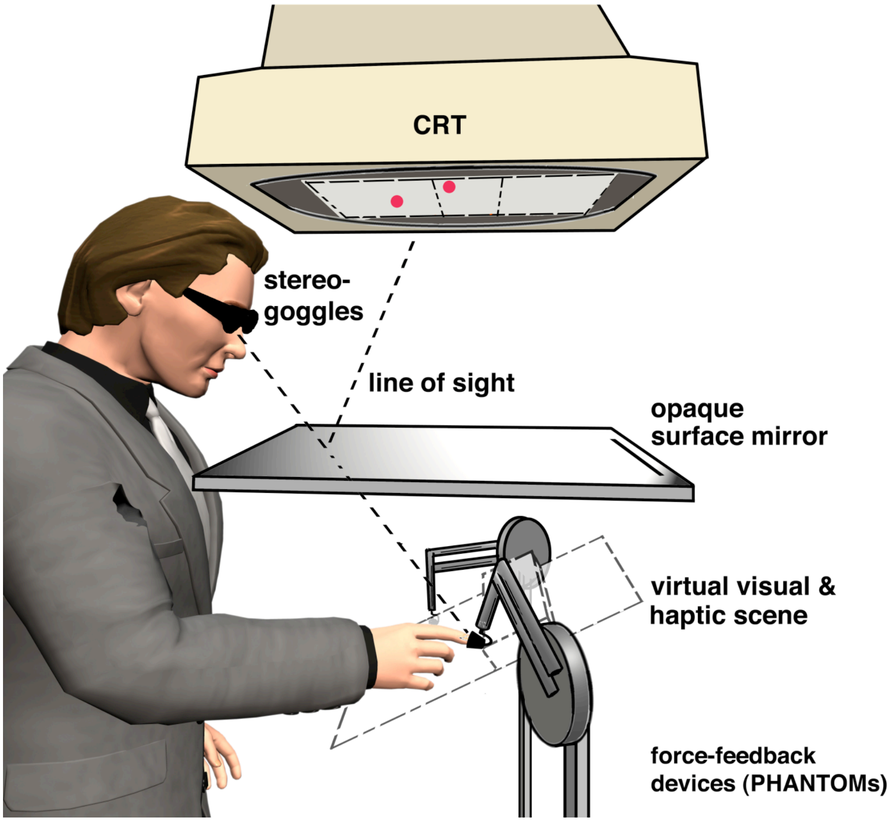


Figure 3: In the visual-haptic setup, observers view the reflection of the visual stimulus binocularly in a mirror using stereo goggles. The haptic stimulus is presented using two PHANToM™ force-feedback devices, one each for the index finger and thumb. With this setup the visual and the haptic virtual scenes can be independently manipulated. *(Taken with permission from Ernst & Bülthoff, 2004)

To determine the reliability of each sensory modality alone, we conducted within-modal visual-only and haptic-only discrimination experiments. The reliabilities can be deduced from the *JND* measurements using Eqs. 2 and 3. The within-modal reliabilities can then be used to come up with predictions for cross-modal performance. Two kinds of predictions can be made that, if confirmed experimentally, would demonstrate optimal integration behaviour. On the one hand we can make predictions for the weights of the signals (using Eqs. 2, 3, and 5); experimentally we can derive the visual and haptic weights from measurements of the Point of Subjective Equality (*PSE*) in a cross-modal experiment. On the other hand we can make predictions for the variance of the combined percept (using Eqs. 2, 3, and 6). Combined variance can be determined experimentally from the cross-modal *JNDs*. It is important to note that these predictions have no free parameter and are merely based on the within-modal *JNDs*.

To determine whether combined visual-haptic performance is statistically optimal according to the MLE model, we again used a 2-IFC discrimination task. In the standard interval we now introduced a small discrepancy between visual and haptic size information ($\Delta = \pm 3$ mm and ± 6 mm). Using the comparison stimulus that was varied in size between 45 mm to 65 mm and that contained no size discrepancy between visual and haptic information we determined the size that was perceived equally in comparison stimulus and standard stimulus (the *PSE*). This perceived size (and so the *PSE*) directly depends on the weights of the individual signals. Given the four different noise levels and the within-modal visual-only and haptic-only discrimination data (JND_V and JND_H) we can calculate the relative visual reliabilities to be 0.78 for 0% noise, 0.75 for 67% noise, 0.48 for 133% noise and 0.16 for 200% noise. These relative visual reliabilities are predictions for the visual weights. As can be seen in Fig. 4, predicted *PSEs* correspond well with the empirically determined *PSEs* ("perceived size" as determined by the

50% point of the psychometric functions). In the no noise condition the *PSE* is close to the visual standard demonstrating a high visual weight. In the 200% noise condition the *PSE* is close to the haptic standard demonstrating a low visual but a high haptic weight. With added noise to the visual display, there is a smooth transition from visual dominance to haptic dominance.

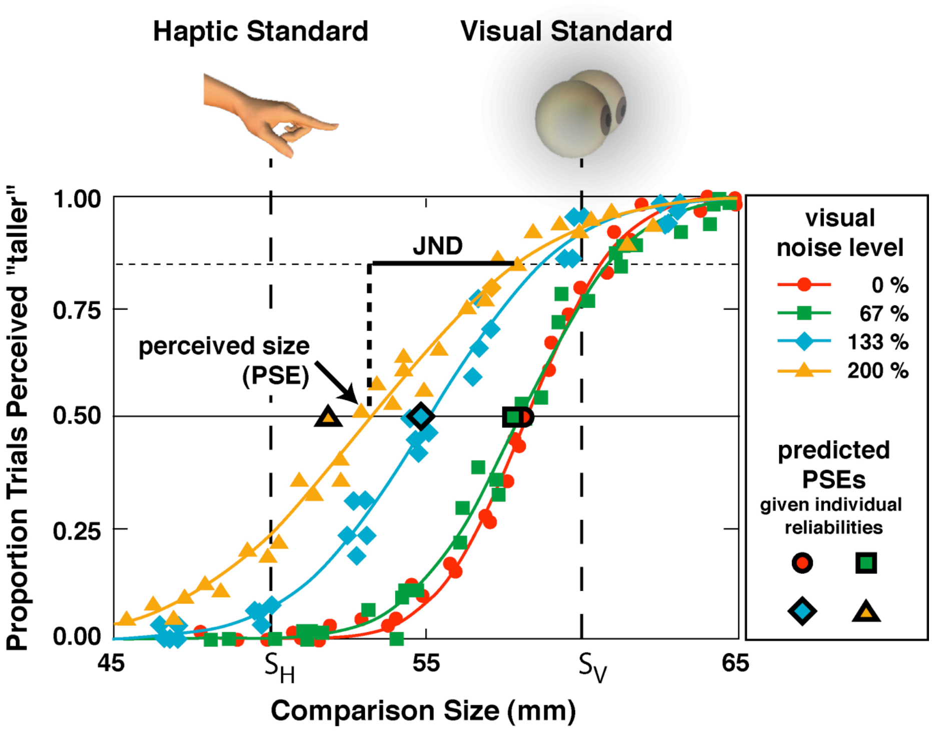


Figure 4: Visual-haptic size-discrimination performance determined with a 2-interval forced-choice task (Ernst & Banks, 2002). We manipulated the relative reliabilities of the individual signals by adding noise to the visual display (0%, 67%, 133%, and 200%). With these different relative reliabilities we measured four discrimination curves. When the relative visual reliability decreases with added noise the perceived size as indicated by the *PSE* is more and more determined by the haptic size estimate (haptic standard) and less by the visual size estimate (visual standard). This demonstrates the weighting behaviour adopted by the brain and the smooth change from visual dominance (red circles) to haptic dominance (orange triangles). As indicated in the figure, the *PSE*s predicted from the individual visual and haptic discrimination performance (symbols with black outline) correspond well with the empirically

determined *PSE*s in the combined visual-haptic discrimination task. Four naïve subjects participated. *(Figure adapted with permission from Ernst & Bülthoff, 2004)

A correct prediction of weights (and *PSE*s) is a first hint that information is combined optimally. However, there are different strategies that would give the same result. For example, this may be a switching strategy in which the observer bases his/her answers on the estimate of one or the other modality at a time but switches answers between the modalities in proportion to their relative reliabilities (Landy & Kojima, 2001; Ernst & Bülthoff, 2004). Even though such a strategy would provide the same weights as are predicted from the MLE model, using such a strategy the combined *JND*s could never become lower than the *JND*s for each individual estimate alone. Therefore, a stronger test of statistical optimality is to show that crossmodal visual-haptic estimates become more reliable when combined; that is, they have a lower *JND*. Predictions for combined visual-haptic *JND*s can be derived from the within-modal *JNDs* using Eq. 6. Experimentally we determined the combined visual-haptic *JNDs* from the cross-modal psychometric functions. As can be seen in Fig. 5, the predicted and empirical *JND*s correspond well. This demonstrates that humans actually combine visual and haptic size information in a fashion that is indistinguishable from statistical optimality (Ernst & Banks, 2002).

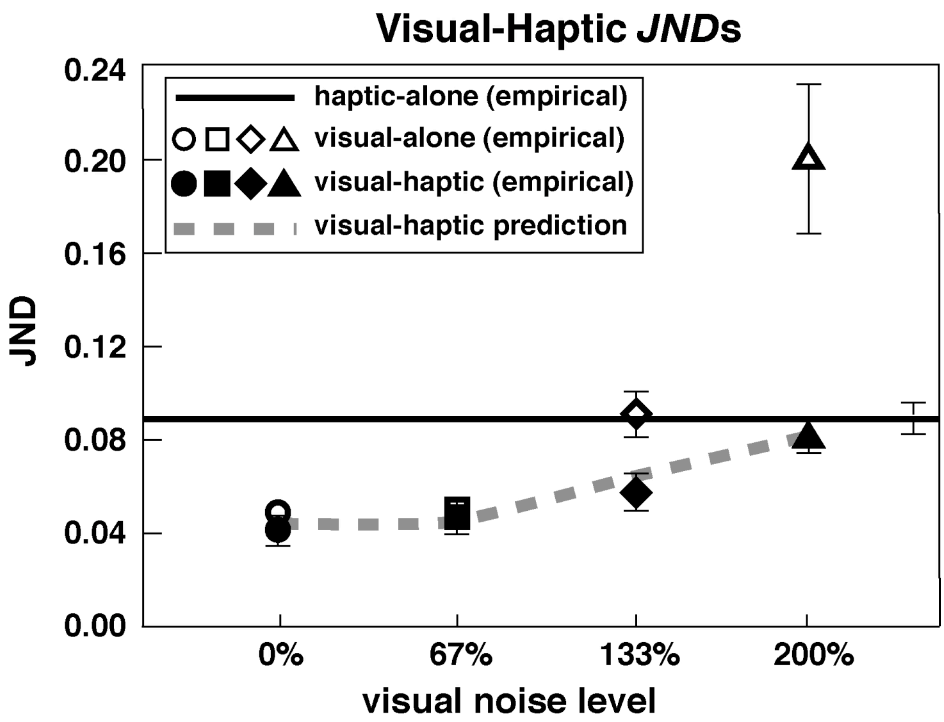


Figure 5: Size discrimination thresholds (*JND*s) for visual-alone, haptic-alone and visual-haptic combined comparisons. For the visual stimulus we used 4 different noise levels (see text). The reliability of the haptic stimulus was not altered. From the individual visual-alone and haptic-alone *JND*s we derived predictions for the combined visual-haptic performance using Eq. 6 (dashed line). Predicted and empirically determined visual-haptic *JND*s correspond well for all four visual noise levels (Ernst & Banks, 2002). Error bars denote the standard error of the mean across subjects (n=4). *(from Ernst & Banks, 2002; with permission)

4 - Benefits and potential costs of integrating information

As demonstrated above, integrating sensory information has the potential benefit of reducing the variance of the associated sensory estimate and increasing its reliability. However, integrating sensory information may also come at a “cost”. The brain potentially may lose access to the individual input signals feeding the integrated percept. That is, by jointly presenting visual and haptic information, as we have done in the experiment presented before, it may be impossible for the brain to independently access the individual sensory information without it being influenced by the other signal. If the brain lost access to the individual estimates, we should be able to observe metameric behaviour. That is, there may be different physical stimuli that lead to exactly the same perceptual experience, indiscriminable from one another. Such metameric behaviour is demonstrated in Fig. 6, right panel.

Fig. 6 is a schematic diagram that shows the expected discrimination performance, first if the two cues are completely independent (left panel), and second if the two cues are fused into a single percept (right panel). With two independent cues, that is the percept of each cue is unbiased when presented in combination, the likelihood functions are radial symmetric in a *JND* normalized cue_1&cue_2-space (for simplicity we here assume that the noise distributions of the signals are Gaussian with constant σ, and that there is no correlation between the noise distributions of the cues; cf. Fig. 11). With that, there is no direction in the cue_1&cue_2-discrimination space that is particularly special. Given an optimal decision rule that takes both independent cues into account (statistical benefit; Graham, 1989), discrimination performance for discriminating a cue_1&cue_2-stimulus from the standard is equal in all directions. Therefore, in the independent-cue case discrimination thresholds around the standard object will form a circle in the cue_1&cue_2-space.

The situation is different if two cues are not independent but instead are fused into a single percept (Hillis, Ernst, Banks, & Landy, 2002). To be optimal the fusion rule is to form the weighted average between the cues (Eq. 4; $cue_{12} = w_1\Delta cue_1 + w_2\Delta cue_2$). That is, if two cues are totally fused at the perceptual level, the same percept will result whether they both indicate a medium value, or if they differ radically from one another but average to a medium value. Hence a high value on one cue can always be compensated for by a low value on the other cue. In other words, fusion is equivalent to averaging a two dimensional stimulus (cue_1–dimension and cue_2–dimension) onto a single, fused dimension. This fused dimension is the positive diagonal in Fig. 6, right panel, where the cue_1 signal maps onto the cue_2 signal. If, for example, cue_1 and cue_2 were visual and haptic sizes, respectively, the positive diagonal is the line where visual and haptic sizes are equal, i.e., it is the common size axis.

If cues are fused it should be obvious that also discrimination performance will be affected. Along the fused dimension (e.g., along the common size axis in the visual-haptic example) discrimination will remain possible. However, for stimuli not being along this dimension discrimination performance will drop and it will be impossible to discriminate stimuli that were averaged to exactly the same value – i.e., that are perceptual metamers. Stimuli that form the same average fall along one particular direction ($\Delta cue_1/\Delta cue_2$) in this discrimination space. This direction is $\Delta cue_1/\Delta cue_2 = -w_2/w_1$ (Hillis, et al., 2002). In Fig. 6 (right panel) where cues are plotted in *JND* units all objects lying on the negative diagonal have the same average and therefore form perceptual metamers with respect to the standard object. Objects lying on lines parallel to the negative diagonal form a different set of perceptual metamers.

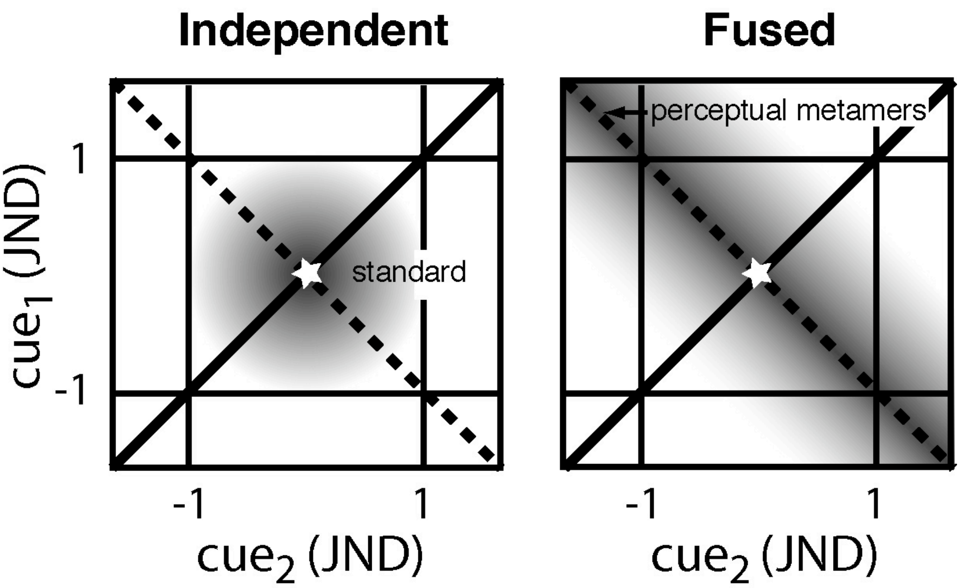


Figure 6: Schematic drawing of predictions for discrimination performance for independent cues (left) and for completely fused cues (right). Plotted is the hypothetical discrimination performance of a two-cue stimulus (cue_1, cue_2) from the standard stimulus (white star; $cue_{1,0}$, $cue_{2,0}$). The cue_1&cue_2-discrimination space is shown in units of *JND*. The grey shading is coding for discrimination performance (dark being non-discriminable to white for perfect discriminability). As indicated, if the two cues (cue_1 and cue_2) are independent there is no explicit direction in this discrimination space and the discrimination contour is a circle around the standard object. If the cues are fused, however, using the weighted averaging rule (see MLE model), objects along the negative diagonal (dashed line in right panel) are not discriminable, i.e., they are perceptual metamers with respect to the standard object.

With this in mind, one can now also imagine that discrimination performance may not fall in one of the two categories illustrated in Fig. 6, but actually may lie somewhere in-between. That is, discrimination performance may not be circular around the standard and it also may not be a single diagonal stripe in discrimination space. Instead, discrimination performance might form an ellipse elongated in the direction that is determined by the weight of the signals ($\Delta cue_1/\Delta cue_2 = -w_2/w_1$) – the same direction that indicates metameric behaviour in the fused case. The magnitude of the elongation of the discrimination ellipse is a reflection of the degree to which the original information feeding into the combination process is accessible and so defines the "*strength of coupling*" between the signals.

We investigated such discrimination performance for two different sets of stimuli using a 3-interval oddity task (Hillis, et al., 2002). In a cross-modal experiment we investigated visual-haptic discrimination performance for object size (Fig. 7, left panel). In a within-modal experiment we investigated slant discrimination performance (Fig. 7, right panel). The slant was defined by two visual cues, one being binocular disparity, the other being texture. Depending on the set of signals used we found different strengths of coupling between the signals, i.e., we found strong coupling for signals derived from within the visual modality (disparity and texture signals to slant) and weaker coupling for visual and haptic signals to size. This can be seen in Fig. 7: in both cases the discrimination thresholds form an elliptical shape – even though in the disparity-texture condition this ellipse is somewhat distorted. This is due to the fact that the reliability of disparity-defined and texture-defined slant changes with slant rather than remaining constant. The MLE model for complete fusion predicts the deformation in the data as can be seen from the green prediction lines (for details see Hillis, et al., 2002). The single cue *JNDs* (red lines), the predicted constraint lines for fused performance (green linens) and the discrimination data can be seen in Fig. 7. The elongation of the discrimination "ellipse" is clearly more pronounced in the disparity-texture experiment than in the visual-haptic experiment. That is, we clearly observed metameric behaviour in the disparity-texture condition, where discrimination performance for the combined stimuli in the upper left and the lower right quadrant of the figure is worse than single cue discrimination performance (single cue *JNDs* are indicated by the red lines). Metameric behaviour is less pronounced in the visual-haptic condition. However, in both conditions performance is not truly metameric over a wide range. In conclusion there seems to be a stronger coupling (interaction) between the within-modal disparity-texture signals than between the cross-modal visual-haptic signals.

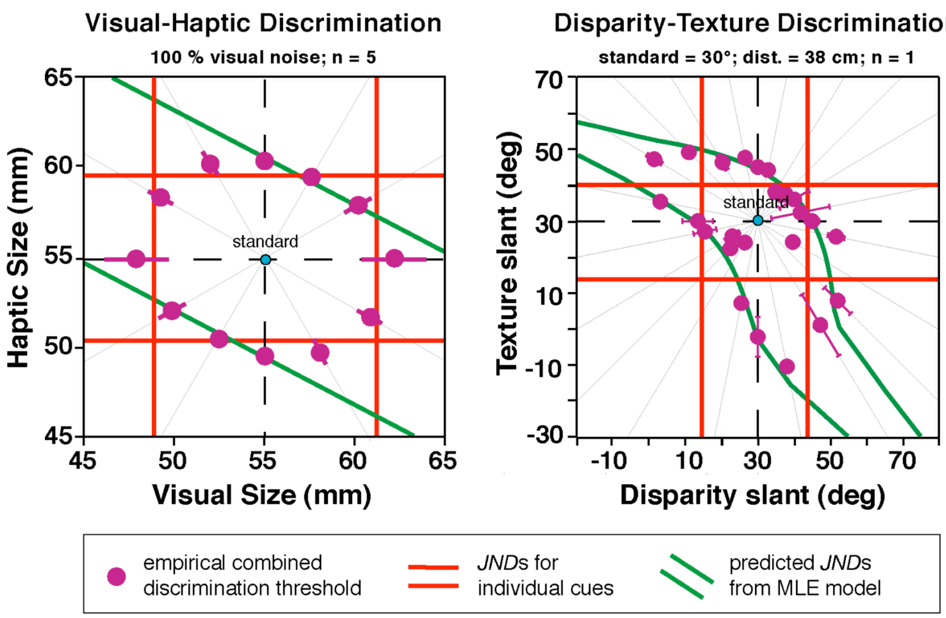


Figure 7: Discrimination performance for visual and haptic size discrimination (left panel) and for binocular disparity and texture slant discrimination (right panel). In red are the single-cue *JND*s for discriminating each individual cue from the standard (green circle). The pairs of red horizontal constraint lines correspond to the *JND*s for haptic size discrimination and texture slant discrimination, respectively. The pairs of red vertical constraint lines correspond to the *JND*s for visual size discrimination and disparity slant discrimination, respectively. In green we plot the discrimination predictions (*JND*s for discrimination from standard) when assuming the MLE model for integration (i.e., complete fusion of the signals). Discrimination performance

in the direction of the green constraint lines indicates metameric behaviour. In purple plotted are the discrimination thresholds in the different directions for combined cue performance (Hillis et al., 2002).

How can the different levels of interaction that determine the strength of coupling between the two sets of cues be understood? As described above, in case of complete fusion the system loses access to the individual estimates. If for some reason, however, the system needs to retain access to the individual estimates, it makes sense that the system does not fuse the signals completely. One obvious candidate for when it is necessary to retain access to the individual estimates is if the mapping between signals is not fixed, but changes in response to a constant conflict between signals. Without some degree of access to the individual estimates it would be impossible to detect conflicts between signals and so it would be impossible for the system to change the mapping between signals in order to overcome the conflict.

Starting with von Helmholtz (1867), there is a huge literature on visuomotor and visual-haptic adaptation, demonstrating the enormously rapid and flexible plasticity of the human visuomotor system (for a review see Welch, 1978; compare also chapter by Holms and Spence in this volume). This clearly indicates that there are often situations in which the mapping between the visual and haptic cues changes. Therefore, in the visual-haptic case it makes sense that the system by and large maintains access to the individual signals.

There is also adaptation for signals within vision. Adams, Banks and van Ee (2001) for example observed an adaptation effect after wearing special magnifying glasses for a number of days and they demonstrated that this adaptation effect is the consequence of recalibration of the relationship between disparity and perceived slant (i.e., a change in mapping). This is reasonable because certainly during growth the relationship between binocular signals, which depend on the interocular distance, and other monocular signals changes and the system needs to adjust for that. However, compared with crossmodal visual-haptic adaptation, which can happen within seconds and a few exposures to the conflict, within visual adaptation seems to be much slower, sometimes taking days or weeks.

The mapping between the different signals can fluctuate on different time-scales. For adapting the mapping between signals quickly, a reliable estimate of the conflict is needed; for adapting slowly, each estimate of the conflict does not need to be very precise – only the average over many observations must yield a reliable conflict estimate. Strong coupling between signals that introduces a strong perceptual bias yields a less reliable estimate of possible conflicts than does weak coupling between signals. If the relationship between signals derived from the same object or event is never changing (the mapping is constant), the system does not need to retain access to the individual estimates and the signals can be completely fused. Because the mapping between the disparity and texture signals does not change quickly (Adams, Banks, & van Ee, 2001), we can infer that for disparity and texture cues to slant it is not so critical to retain reliable access to the individual estimates and therefore there can be strong coupling between the disparity-texture signals. The reverse is true for visual-haptic size signals.

In conclusion, it seems that the necessity for changes in the mapping and how quickly they should occur determines the *strength of coupling*. In the following section I aim to explain this *strength of coupling* using a Bayesian prior.

5 - The "*Strength of Coupling*" and Bayesian Decision Theory

In the previous sections I have demonstrated that humans integrate visual and haptic information in a statistically optimal fashion. I have also shown that complete fusion between visual and haptic signals does not occur; rather there seemed to be only some weaker coupling between the signals. Is that a contradiction or is there a common explanation? I will try to answer this question using Bayesian Decision Theory (BDT).

To better understand the integration mechanisms it is useful to examine more closely the *strength of coupling* between the sensory signals. Therefore, I present next an experiment that directly analyses the extent to which the individual signals are accessible when provided with a combined visual-haptic stimulus. Determining the accessibility of the individual signals in combination can be done using a mixed design in which a combined visual-haptic stimulus is compared to either a visual-only or a haptic-only stimulus. When the comparison stimulus is visual-only subjects are instructed to ignore the haptic component of the combined stimulus and vice versa. Visual-only and haptic-only trials were randomly intermixed. By introducing a small discrepancy between the visual and haptic information in the combined stimulus, one can estimate which signal has the higher weight. If the brain retains access to the individual visual and haptic information then when compared to the visual–only stimulus there should be no influence of the haptic modality (i.e., a visual weight of one) and vice versa (i.e., a visual weight of zero when the combined stimulus is compared to the haptic-only stimulus). On the other hand if in combination visual and haptic information is completely fused (i.e., not accessible independently), comparing the combined stimulus to either the visual-only or the haptic-only stimulus should provide identical results (i.e., the relative weight of the signals should be identical in both conditions).

The results of such an experiment can be seen in Fig. 8 (Ernst & Banks, 2000). For this experiment we used a 2-IFC task to measure discrimination performance. Subjects compared a combined visual-haptic stimulus to either a visual-only stimulus (solid line) or a haptic-only stimulus (dashed line). We determined the relative visual weights from the *PSE*s of these psychometric functions (for details of the methods see Ernst & Banks, 2002). For each of the two conditions we derived weights for four different reliability levels. As in the previous experiment (Fig. 4) we varied the reliability of the signals by adding noise to the random-dot pattern that constitutes the visual display (0%, 67%, 133%, and 200% visual noise). The stimulus was exactly the same as described above and by Ernst and Banks (2002).

Independence of signals predicts a visual weight of zero if compared to the haptic-only (H) stimulus and a visual weight of one when compared to the visual-only (V) stimulus. Complete fusion of signals predicts that the visual weights in the two conditions should be identical. Neither prediction was confirmed. Instead we found relative visual weights that were in-between these two predictions. The relative visual weights differed between the two conditions. Vision was weighted more heavily when the combined stimulus was compared to the visual modality and the haptic modality was weighted more heavily when the combined stimulus was compared to the haptic modality. Such a result was obtained across all four noise-levels (Fig. 8). That is, it seems that we have no direct access to the original haptic or visual information feeding into the combination process, but instead this information is altered (biased) by the accompanying modality. In other words, subject cannot ignore the task-irrelevant stimulus even when they are instructed to fully attend to only one of the two sensory signals and to ignore the other. It is probably worth noting that there are several studies now showing that this form of sensory integration is not modulated by attention (e.g., Bertelson, Vroomen, De Gelder, & Driver, 2000; Vroomen, Bertelson, & De Gelder, 2001).

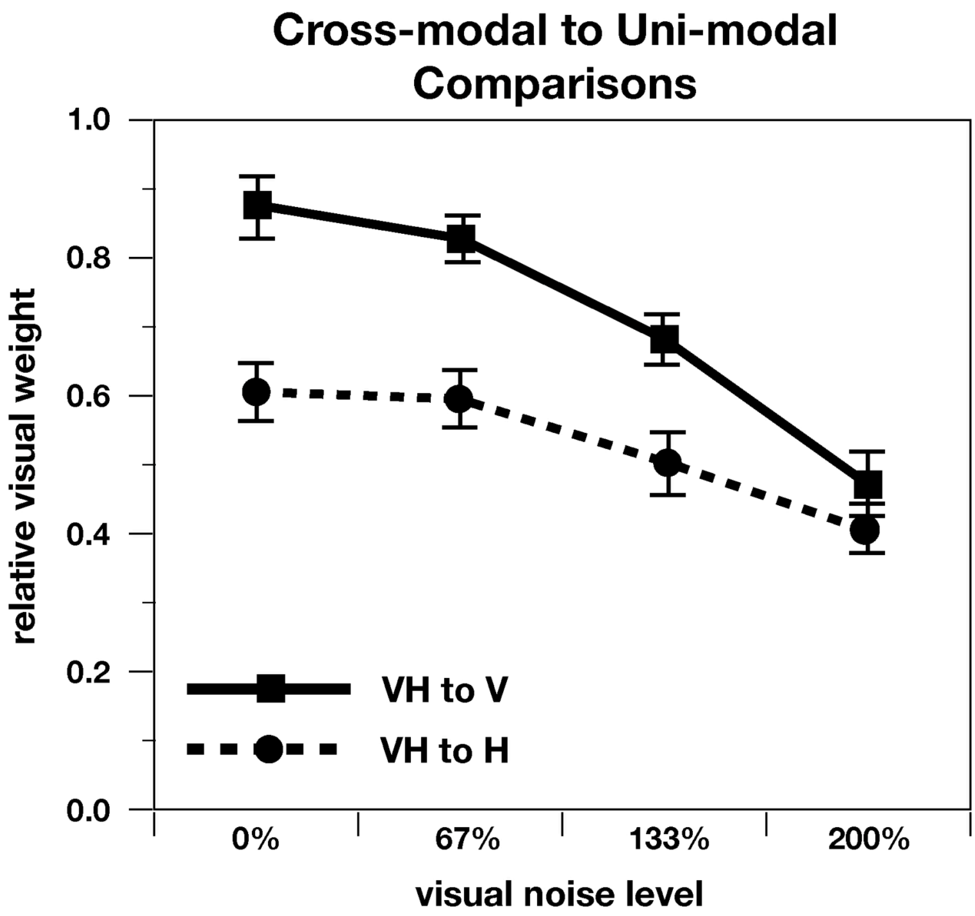


Figure 8: Cross-modal to uni-modal comparisons. The weights are determined using a 2-IFC size-discrimination task with a stimulus similar to Ernst and Banks (2002). The solid line indicates visual weights that were determined from a condition where the combined visual-haptic stimulus (VH) was compared to a visual-only stimulus (V). The dashed line indicates visual weights that were obtained from a condition where the combined visual-haptic stimulus (VH) was compared to a haptic-only stimulus (H). Trials from both conditions were randomly intermixed. We determined weights in both conditions for four different noise levels to alter the relative reliability of the signals. The noise was added to the visual display only. Error bars denote the standard error of the mean across subjects (n=4).

If the visual-haptic signals are not completely fused into a unified percept and they are also not independent, in which case the percept would be unbiased, what is the percept that is associated with the visual-haptic stimulus? From the VH-to-V experiment we can conclude how much the visual size percept is biased by the presence of the haptic component in the combined visual-haptic stimulus. Conversely, from the VH-to-H experiment we can conclude how much the haptic size percept is biased by the presence of the visual component in the combined visual-haptic stimulus. The visual and haptic biases correspond to the weights (*PSEs*) determined in the VH-to-V and VH-to-H experiments, respectively. No visual bias would mean that visual weight is $w_V=1$ in the VH-to-V experiment; no haptic bias would mean that visual weight is $w_V=0$ in the VH-to-H experiment (independence). As indicated previously, a bias in the percept that would indicate complete fusion would result in identical weights in the two experiments (VH-to-V and VH-to-H).

To illustrate the percept associated with a visual-haptic stimulus the results from Fig. 8 are re-plotted in Fig. 9 in a 2-dimensional visual-haptic space. Haptic size is plotted on the ordinate, visual size on the abscissa. In general, a visual-haptic stimulus (unfilled circle) results in a size percept (filled circle) that is biased in vision and in touch relative to the physical size stimulus ($S_{V,0}$, $S_{H,0}$). Quantifying these biases (weights) requires a visual-haptic stimulus that is off the identity line. The identity line for which the visual and haptic sizes are equal is the positive diagonal in Fig. 9.

The bias corresponds to a relative size shift between stimulus and percept ΔS and depends on the weights. The visual size shift with weights determined in the VH-to-V experiment corresponds to

$$\Delta S_V = (1 - w_V) \cdot (S_{V,0} - S_{H,0}); \tag{7}$$

the haptic size shift with weights determined in the VH-to-H experiment corresponds to

$$\Delta S_H = w_V \cdot (S_{V,0} - S_{H,0}). \tag{8}$$

In Fig. 9 I plot these biases for the four visual noise conditions: on the abscissa there is the visual size shift, on the ordinate the haptic size shift. The dashed arrow indicates the overall bias from physical to perceived size of the visual-haptic stimulus. The length of this arrow (in relation to the distance between stimulus and identity line) is an indication for the *strength of coupling* between the visual and haptic signals. The more pronounced the bias, the longer the arrow, the stronger the coupling. If the estimates were independent, perceived size would be identical to physical size and so the length of the arrow would be zero. If the visual and haptic signals are fused completely the percept will lie on the identity line (positive diagonal) corresponding to a maximal bias and the maximal possible length of the arrow. As now can also be seen in Fig. 9, the actual results fall in-between independence and complete fusion.

The direction of the bias (orientation of the arrow) should correspond to the relative reliabilities of the visual and haptic signals, i.e., the optimal combined visual-haptic weights. If the visual information is more reliable than the haptic information there is a stronger haptic than visual bias, so that the orientation of the arrow is closer to horizontal (Fig. 9 upper two panels). Conversely, if the haptic information is more reliable than the visual information there is a stronger visual than haptic bias, so that the orientation of the arrow is closer to vertical (Fig. 9, lower right panel). Equal reliabilities of the signals should correspond to a bias along the negative diagonal (Fig. 9, lower left panel). As can be seen in Fig. 9 the directions of the biases for all different noise levels correspond well to the relative reliabilities of the visual and haptic signals (cf. Fig. 5); that is, the direction of bias corresponds well to the optimal combined visual–haptic weights.

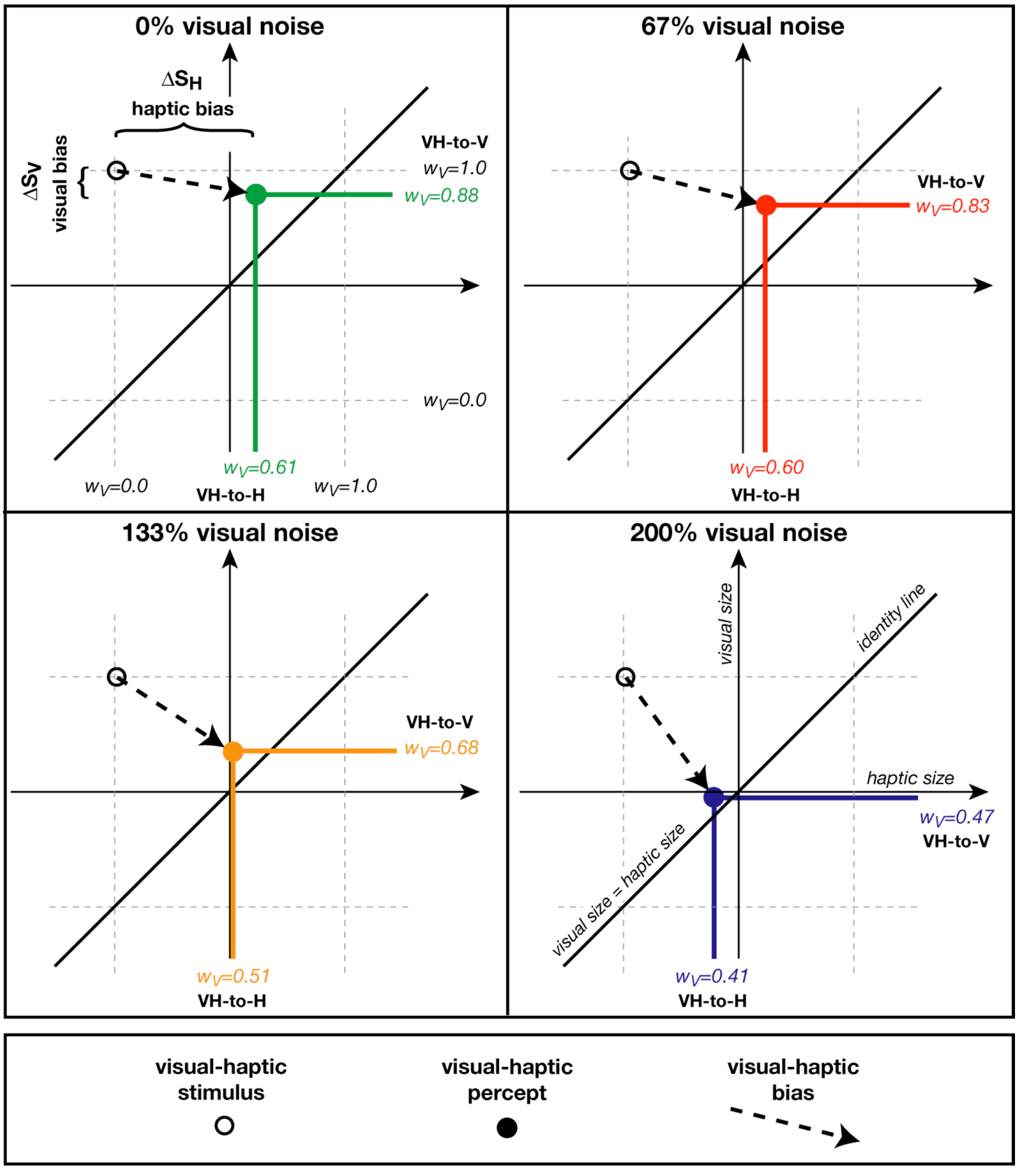


Figure 9: Visual-haptic percept (filled circles) in relation to the stimulus (unfilled circles) that gave rise to it. The four panels show the results for four different noise levels. The abscissa indicates the visual size, the ordinate the haptic size of the object. Visual sizes were compared in the VH-to-V condition, haptic sizes in the VH-to-H condition. Form this the visually and haptically perceived sizes were determined using *PSE*s (weights). The difference between physical and perceived size (the visual-haptic bias; dashed arrow) directly depends on the weights of the signals (Eqs. 8 and 9; see text).

To bring all the findings discussed above together Bayesian Decision Theory (BDT) seems to be an appropriate common framework (Fig. 10). Bayesian Inference provides a formal way to model uncertainty about the world by combining prior knowledge (the prior) with observational, sensory evidence (the likelihood function) to infer the most probable interpretation of the environment (the posterior) (Yuille & Bülthoff, 1996; Knill & Richards, 1996; Mamassian, Landy, & Maloney, 2002; Kersten & Yuille, 2003). Bayes' Rule states that the posterior probability $p(W \mid I)$ is proportional to the product of the likelihood function $p(I \mid W)$ and prior probability distribution $p(W)$: $p(W \mid I) \propto p(I \mid W) \times p(W)$. In general, the Bayesian Framework can be used to construct 'ideal observer' models as a standard for comparison with human performance. This framework has recently seen much success in describing observers' perception and performance in a variety of visual (Knill, 1998; Saunders & Knill, 2001; Weiss, Simoncelli & Adelson, 2002; Adams & Mamassian, 2004), visual-haptic (Ernst & Banks, 2002, Hillis, et al., 2002, Adams, Graf, & Ernst, 2004) visual-auditory (Alais, & Burr, 2004), and visuomotor coordination tasks (Körding, & Wolpert, 2004).

The first step in BDT is to construct the posterior. After combining the prior and the likelihood function into the posterior distribution using Bayes Rule, to perform an action or to come to a decision, the second step is to define the goal for the task using gain/loss functions (Fig. 10) (Schrater & Kersten, 2000; Mamassian, et al., 2002). That humans can behave very close to optimal when making decisions or actions was recently demonstrated by Trommershäuser, Maloney, and Landy (2003). They showed that statistical decision theory can be used to accurately explain pointing behaviour under risk using different cost functions. A complete Bayesian model has to consider all three parts that make up Bayesian' Decision Theory: sensory estimation, prior knowledge, and a decision-making process (e.g., Kersten, 1999; Mamassian, et al., 2002; Ernst & Bülthoff, 2004).

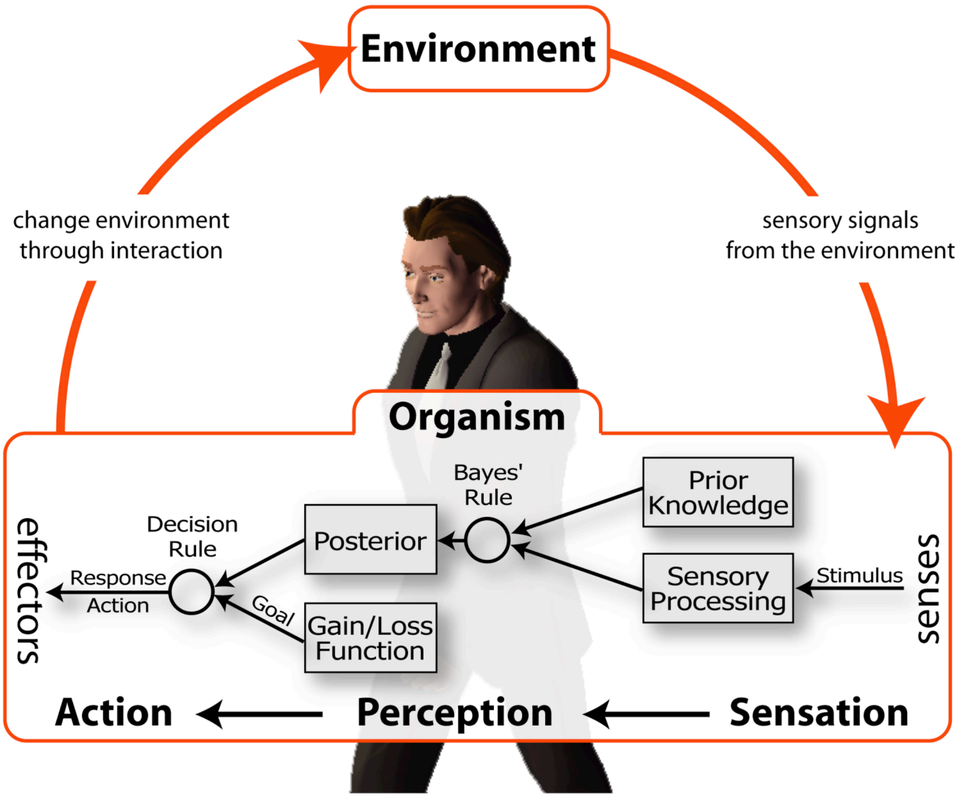


Figure 10: Perception/Action-Loop including Bayesian Decision Theory (BDT). See text for details. *(taken with permission from Ernst & Bülthoff, 2004)

To model multimodal cue integration using a Bayesian approach, a prior is necessary describing the interactions between the signals. A robust system that behaves plastically to body changes or to modifications in the environment often has to adapt the mapping between its signals. As suggested above, the ease for adapting the mapping depends on the *strength of coupling* between the signals (i.e., their interaction): weak coupling allows for more vigorous changes than does strong coupling. However, not all mappings between two signals are equally likely. For example, if the signals are visual and haptic sizes the changes in the mapping that naturally occur are in the order of a few millimetres only. As a result there is a probability distribution representing the mappings naturally occurring. In the following I will show that a Bayesian prior that corresponds to the probability distribution for the different mappings between two signals can be used to model the observed sensory interactions.

Such a prior that codes for the mapping between two signals is aligned along the identity line. In the simplest case spread of the prior will be Gaussian distributed (see Fig. 11, middle column). The standard deviation of this distribution determines the influence of the prior. That is, the sensory signal (the likelihood function) is biased more if the prior is represented by a narrow (Fig. 11, C-G) than a wide (Fig. 11, B) distribution. If the prior is completely flat, that is, the standard deviation approaches infinity, the posterior is equal to the likelihood distribution and the bias is negligible (Fig. 11, A).

The likelihood function associated with a sensory signal represents the sensory information available to the system. It is determined by a mean and its variance (Fig. 11, left column). The smaller the variance associated with a sensory signal the more reliable is the sensory information. The posterior, which is proportional to the product of likelihood and prior, determines the percept (the *Maximum a Posteriori* or MAP estimate). That is, both the likelihood and the prior affect the percept. The strength of the bias depends on the relation between the two distributions. If the likelihood is very reliable it can be less biased by a prior. Vice versa, if the prior is given by a narrowly tuned distribution it will bias the likelihood more. If the prior is reduced to a delta-function (variance approaches zero) the Bayesian scheme discussed here corresponds to the MLE model discussed earlier in this chapter.

The influence of the prior therefore determines the *strength of coupling* between the signals and the degree to which the signals interact. Hence, we call this prior the "*Coupling Prior*" and it relates to the probability for knowing the mapping between sensory signals. If the mapping is known for sure, signals can be fused; contrary, if the mapping is unknown, signals should be kept separate.

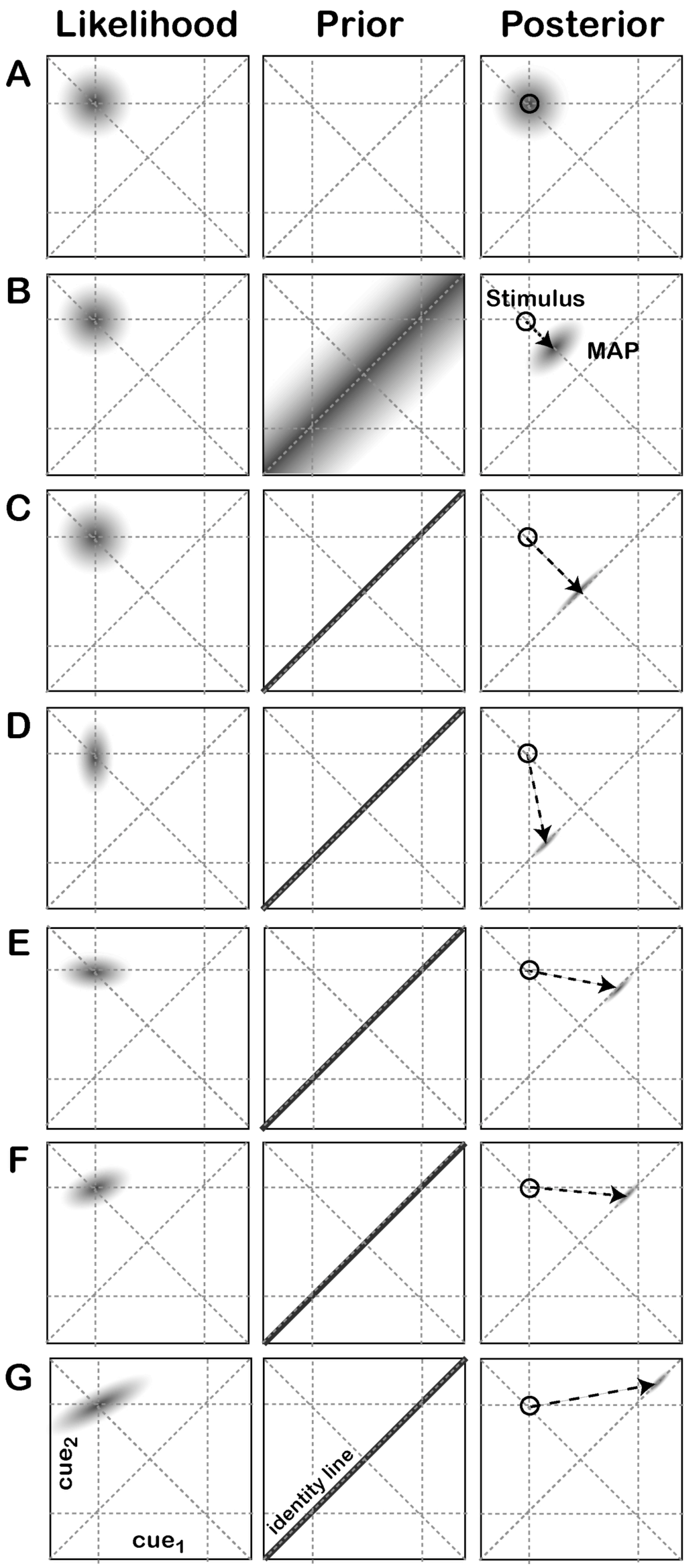

Likelihood
Prior
Posterior
A
B
Stimulus
MAP
C
D
E
F
G
cue2
cue1
identity line

Figure 11: Schematic illustration of several examples demonstrating sensory combination with a Bayesian Prior. In the left column there are the likelihood distributions indicating the sensory information available. In the middle column are the priors that relate to the probability distributions for the mappings between the sensory signals. The multimodal percept is based on the posterior (right column). According to Bayes' rule the posterior is proportional to the product of likelihood and prior. The relationship between likelihood and prior, therefore, determines the degree of coupling between the signals. A) The prior is flat so that the likelihood equals the posterior. This indicates independence between the signals. B) The prior is aligned along the identity line and is moderately spread. The cues therefore show some moderate interaction (coupling). C) Prior is tuned very sharply (delta-function). This situation corresponds to complete fusion. D) and E) Only the relative variances of the two signal estimates determine the direction of bias. This corresponds to the weighting of the signals. (C, D, and E correspond to the MLE model discussed earlier in this chapter). F) In all cases discussed previously it was assumed that the noise distributions of the two signals are independent. If they are correlated the likelihood ellipse is rotated in the cue_1&cue_2-space. This slightly changes the weighting between the signals (see Oruç, et al., 2003). G) If there is a positive correlation (ρ_{12}) between the noise distributions of the cues and the reliabilities differ substantially so that $\rho_{12} > \sqrt{r_1/r_2}$, the weight for cue_1 will become negative and the weight for cue_2 bigger than 1 (the sum of the weights must be 1) (see Oruç, et al., 2003). In this example the direction of bias is therefore above vertical.

Given the priors used in the examples shown in Fig. 11 (all aligned along the identity line), the direction of the bias is only determined by the relative reliabilities between the two sensory signals (cues). The relative reliabilities of the signals therefore define the direction in which the perceptual estimate is biased by the *Coupling Prior* (Fig. 11, C-E). In the visual-haptic example discussed above, if the visual estimate is more reliable than the haptic estimate the bias is closer to the visual modality and vice versa. This corresponds to the weighting of the sensory signals. Thus, both the weighting of the signals, i.e. the direction in which the percept is biased by the prior, as well as and the strength of coupling between the signals can be explained using this Bayesian approach.

In the experiment shown in Fig. 9, the strength of coupling increased slightly with added visual noise in the four noise conditions (i.e., the more noise is added the closer the combined percept is to the identity line). This effect can also be explained with the Bayesian framework introduced. The variance of the associated sensory estimate (likelihood function) increases with added noise. Therefore, with a likelihood function that is less salient the *Coupling Prior* can become more influential.

Introducing the *Coupling Prior* can thus explain the results of the last two experiments that were presented in Figs. 7 and 8. These experiments demonstrated that there is no complete fusion but some weaker coupling between the sensory signals. In the first experiment (Figs. 4 and 5), however, we found that subjects integrate information in a statistically optimal fashion. Does this mean there was complete fusion in this first experiment? The combined visual-haptic stimulus was identical in all these three experiments. Therefore, it can be assumed that the percept of this stimulus was so as well. The only difference between these experiments was the decision that had to be made with respect to the combined stimulus. In Experiment 1 we used a 2-IFC task in which the combined visual-haptic stimulus was compared to different versions of the same visual-haptic stimulus and the subjects' task was to discriminate size. In Experiment 2 we used a 3-interval oddity task, in which each interval contained a visual-haptic stimulus. Two of the intervals were identical and the stimulus in the odd interval differed in either visual and/or haptic size. The subjects' task was to discriminate the odd visual-haptic stimulus. Experiment 3 was virtually identical to Experiment 1 using again a 2-IFC task. The only difference was that subjects were to do size discrimination between a visual-haptic stimulus and a visual-only or haptic-only comparison stimulus. The difference between the experiments is the task – i.e., the decision or action to be taken upon the identical visual-haptic percept. Therefore, the apparent different experimental results (optimal integration vs. incomplete fusion) seem to be related to differences in the task at the decisional level and not to the differences in the percept of the visual-haptic stimulus per se.

Exploiting BDT these differences in experimental results can be subsumed in the gain/loss functions (cf. Fig. 10, second step). If the goal of the perceptual system is to come up with the optimal (most reliable) decision, it is not necessary that the signals have to be completely fused. In contrast, the system may perform optimal because the decision process is optimal and takes the reliability of the individual sensory estimates into account (not implying that the decisions or actions have to be conscious). Using BDT the decision (action) process can be modelled using gain/loss functions. Gain/loss functions define the goal of the task and thus are task dependent.

Using a 2-IFC task as in the first experiment it cannot be distinguished whether the sensory signals are optimally integrated into a fused percept or whether subjects come up with an optimal decision that integrates the information available taking the reliabilities of the individual sensory signals into account. In both cases subjects' answers would have the lowest variance possible. Integration of signals (meaning optimal use of sensory information) could be performed on either of these two levels – the perceptual level describing the coupling between the signals (task independent) and the decisional level comprising the goal of the task (task dependent). If the signals were completely (mandatorily) fused the percept should not be separable any more by using another task. Residuals of the individual visual and haptic signals, however, were found in Exp. 2 and 3. Here we found incomplete fusion. Thus, to combine the findings of all three experiments, I here assume that multisensory cue combination is a two-step process: In the fist step, the signals are perceptually coupled. This coupling is task independent and can be described with the *Coupling Prior*. In the second step the goal of the task is defined. This goal is defined using gain/loss functions. This second step may also be optimal. That is, even holding a weak *Coupling Prior* performance can still be optimal. Therefore, BDT with an optimal decision stage together with a *Coupling Prior* unites all of the experimental findings presented. Thus, Bayesian decision theory constitutes a comprehensive framework for characterizing sensory integration.

6 - Concluding remarks

Multiple sensory signals derived from several modalities can provide redundant sensory information about one's own body and the environment. This chapter has been a review of mechanisms that may be exploited by the perceptual system to integrate such redundant sensory information into a coherent percept in order to come to an optimal decision or action.

Three experiments on visual-haptic combination of size information were discussed. The first experiment demonstrated that humans integrate visual and haptic size information in a statistically optimal fashion by maximally reducing the variance of the final perceptual decision. However, as demonstrated by the results of the second experiment, this does not necessarily imply that the sensory size signals are mandatory fused into a unified percept. Instead, we found a weaker form of interaction between the size information from vision and touch. The degree of interaction between the sensory signals was taken as a definition for the *strength of coupling* between the signals. We indirectly determined the actually perceived visual and haptic sizes associated with the visual-haptic stimulus in the third experiment. The perceived size of the visual-haptic stimulus was in fact in-between the sizes predicted from either complete fusion or independence of the signals.

Bayesian Decision Theory offers an excellent basis for modelling these findings using a common framework. To account for the sensory interactions, i.e. the coupling between the signals, a Bayesian prior was introduced. This *Coupling Prior* represents the probability distribution of naturally occurring mappings between the sensory signals, i.e. it codes for the certainty of knowing the mapping between signals. If the mapping almost never changes the prior is represented by a very narrow distribution and signals can be fused completely. If the mapping constantly changes, however, the distribution is more spread and signals are fused less vigorously, but may still interact. If there is no mapping between signals (they carry no redundant information), the *Coupling Prior* will be flat and the signals do not interact; i.e., the signals will be independent.

Why does the mapping between signals change? In order to be robust against changes in the environment or changes occurring to the body our perceptual and perceptual-motor system has to be very flexible. The system can compensate for such changes through the process of adaptation, which corresponds to a change in mapping between signals. For example, human beings are very skilled at using tools. Using a tool often requires a mapping from visual coordinates to the coordinates of the tools end-effector. Thus, using tools can be seen as an extension of the body that requires adaptation (see chapter by Làdavas & Famè in this volume). Fusing signals mandatory so that the system loses access to the incoming information would prevent such adaptation from happening, because the discrepancy between the signals could not be detected. Without the discrepancy being detected, the error signal necessary for adaptation (the remapping of signals) is missing. A reliable error signal allows for quick adaptation; adaptation to an unreliable error signal should be slow. Therefore, whenever quick adaptation is necessary (e.g., for remapping of vision to body sense during tool use – plasticity of peripersonal space) signals should not be tightly coupled maintaining the chance to reliably detect an error signal; contrary, when adaptation can be on a longer time frame (as argued for the disparity-texture example from Exp. 2) coupling can be stronger and so the signals are fused more completely.

Integration of signals is only reasonable if they are derived from the same object or event; unrelated signals should be kept separate. For example, we have recently shown that sensory integration breaks with signals that are not in temporal synchrony (Bresciani, Ernst, Drewing, Bouyer, Maury, & Kheddar, 2004) or that come from largely different locations in space

(Gepshtein, Burge, Banks, & Ernst, 2004). In order to know which signals derived from the different modalities belong together the correspondence problem has to be solved. Signals are likely to originate from the same object or event if they are derived roughly at the same time, from roughly the same location in space. The likelihood of correspondence increases with the number of concordant attributes, where each attribute is one dimension in a hyper-dimensional multimodal space. The *Coupling Prior* that is responsible for integration to occur is maximally tuned at the origin when all multisensory attributes agree. If one or more attributes contain a discrepancy between the sensory signals, e.g., when signals are derived at different locations but at the same time, the *Coupling Prior* should become more flat. With a completely flat prior there is no interaction and the signals are independent. This concept of correspondence would therefore predict that integration of sensory signals only occurs when the conflict between sensory attributes is small. As an example consider the integration of visual and somatosensory information. Such signals concerning the spatial position of the body should only be integrated if not in conflict. That is, visual-somatosensory interactions should decrease (integration should break) outside the peripersonal space when visual-somatosensory spatial conflicts become large. Experiments conducted by Làdavas and Famè confirm this prediction (cf. chapter by Làdavas & Famè in this volume).

In the beginning of this chapter I introduced an integration model based on the Maximum-Likelihood-Estimate (MLE). Later I then switched to a Bayesian model that combines the likelihood with a prior. How do those two models relate? The MLE model was formulated in a one-dimensional space (e.g. the space of physical size) and it can describe sensory integration in this one-dimensional space. Once we realized that the visual and haptic signals are not completely fused, the problem of sensory integration became two-dimensional (e.g., visual-size being one, haptic-size the other dimension). For describing such a two dimensional situation the MLE model was not appropriate anymore. I therefore switched to the Bayesian model in which the prior is responsible for the interactions occurring between the sensory signals. Independence could be modelled with a flat prior (the standard deviation σ of the Gaussian prior approaches infinity), complete fusion with a prior that corresponds to a delta function (the standard deviation σ of the Gaussian prior approaches zero). Complete fusion corresponds to a reduction from two to one dimensions and can be described with a prior representing a delta function. Therefore, the MLE model corresponds to the Bayesian model with a *Coupling Prior* that is a delta function (Fig. 11, C-E). In general, compared to the MLE model, the Bayesian model is more comprehensive because using a prior with an intermediate standard deviation σ it can also describe interactions that are in-between complete fusion and independence.

We have shown that sensory signals are not necessarily completely fused into a unified percept, yet subjects can perform in an close to optimal fashion when asked to report the combined percept (Experiment 1). In order to perform optimally under such conditions the nervous system has to know the reliability of the sensory signals and take them into account when making decisions or actions. Using BDT this implies that both stages – the sensory combination stage exploiting the Bayesian prior, as well as the decision stage has to be optimal. Naturally, this does not necessarily imply that the actor is consciously aware of his/her decisions or actions.

As a conclusion, I have proposed that a Bayesian model that uses a *Coupling Prior* for describing sensory interactions is a convenient theoretical framework for understanding multimodal cue integration as a continuous process between independence and complete fusion.

Acknowledgement: This work was supported by the Max-Planck Society and by the 5th Framework IST Program of the EU (IST-2001-38040, TOUCH-HapSys). Thanks to Michael Landy, Knut Drewing and Johannes Burge for helpful comments on this manuscript.

References:

Adams, W. J., Banks, M. S., & van Ee, R. (2001, Nov.). Adaptation to three-dimensional distortions in human vision. *Nature Neuroscience, 4 (11),* 1063-1064.

Adams, W. J., Graf, E. W., & Ernst, M. O. (Oct. 2004) Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7 (10).*

Adams, W. J., & Mamassian, P. (2004). Bayesian combination of ambiguous shape cues. *Journal of Vision* (in press).

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

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics, 62,* 321-332.

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 *J. Neurophysiol. 81 (3),* 1355-1364.

Bresciani, J-P, Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., & Kheddar, A. (2004). Feeling what you hear: auditory signals can modulate tactile tap perception. *Experimental Brain Research (submitted).*

Clark, J. J. & Yuille, A. L. (1990). *Data Fusion for Sensory Information Processing Systems*. Kluwer Academic Publishers.

Cochran, W. G (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society (Supplement), 4,* 102-118.

Ernst, M. O. & Banks, M. S. (2000). What determines dominance of vision over haptics? *Proceedings of the Annual Psychonomics Meeting*.

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 (4),* 162-169.

Gepshtein S. & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology 13*, 483-488.

Gepshtein, S., Burge, J., Banks, M. S, & Ernst, M. O. (2004). Optimal Combination of Vision and Touch has a Limited Spatial Range. *Current Biology (submitted).*

Graham, N. (1989). *Visual Pattern Analyzers*. New York: Oxford University Press.

Green, D. M., & Swets, J. A. (1988). *Signal Detection Theory and Psychophysics*. Peninsula Publishing, Las Altos, California, USA.

Helmholtz, H. von (1867). *Handbuch der physiologischen Optic, Vol III* (Leopold Voss, Leipzig).

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: Effects of Base Slant and Distance. *(in preparation)*.

Kersten, D (1999). High-level vision as statistical inference. In M.S. Gazzaniga (Ed.), *The New Cognitive Neurosciences*. Cambridge, MA, MIT Press.

Kersten D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annu. Rev. Psychol. 44,* 271-304.

Kersten, D. & Yuille, A. L (2003). Bayesian models of object perception. *Curr. Opin. Neurobiol. 13*, 150-158.

Knill, D. C. (1998). Discriminating surface slant from texture: comparing human and ideal observers. *Vision Research, 38,* 1683-1711.

Knill, D. & Richards. W. (1996). *Perception as Bayesian Inference*. Cambridge University Press

Knill, D. C. & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43 (24),* 2539-2558.

Körding, K. P. & Wolpert, D. M. (2004). Bayesian integration of sensorimotor learning. *Nature, 427,* 244-247.

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

Landy, M. S, & Kojima, H. (2001). Ideal Cue Combination for Localizing Texture Defined Edges. *J. Opt. Soc. Am. A 18,* 2307-2320

Mamassian, P., Landy, M., & Maloney, L. T. (2002). Bayesian Modelling of Visual Perception. In *Probabilistic Models of the Brain.* (Eds. Rao, P. N., Olshausen, B. A., & Lewicki, M. S.) MIT Press, Cambridge, MA.

Oruç, I., Maloney, T. M., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43,* 2451-2468.

Saunders, J. A. & Knill, D. C. (2001). Perception of 3D surface orientation from skew symmetry. *Vision Research, 41,* 3163-3183.

Schrater, P. R., & Kersten D., (2000). How optimal depth cue integration depends on the task. *Intern. J. Comp. Vis. 40 (1),* 71-89.

Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics, 63,* 651-659.

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: Adapting to altered sensory environments* (Academic Press, NY).

Yuille, A. L. & Bülthoff, H. H (1996). Bayesian theory and psychophysics. In *Perception as Bayesian Inference,* (Eds.) D. Knill, W. Richards. Cambridge University Press, 123-161.