Reliability and Consistency Affect the Integration of Visual Depth Cues

Inauguraldissertation
zur Erlangung des Akademischen Grades
eines Dr. phil.,

vorgelegt dem
Fachbereich 02 – Sozialwissenschaften, Medien und Sport
der Johannes Gutenberg-Universität
Mainz

von
Matthias Gamer
aus Alfeld/Leine

Mainz
2008

Tag des Prüfungskolloquiums: 4. August 2008

Contents

Appendix

1. Abstract

Visual perception relies on a two-dimensional projection of the viewed scene on the retinas of both eyes. Thus, visual depth has to be reconstructed from a number of different cues that are subsequently integrated to obtain robust depth percepts. Existing models of sensory integration are mainly based on the reliabilities of individual cues and disregard potential cue interactions. In the current study, an extended Bayesian model is proposed that takes into account both cue reliability and consistency. Four experiments were carried out to test this model's predictions. Observers had to judge visual displays of hemi-cylinders with an elliptical cross section, which were constructed to allow for an orthogonal variation of several competing depth cues. In Experiment 1 and 2, observers estimated the cylinder's depth as defined by shading, texture, and motion gradients. The degree of consistency among these cues was systematically varied. It turned out that the extended Bayesian model provided a better fit to the empirical data compared to the traditional model which disregards covariations among cues. To circumvent the potentially problematic assessment of single-cue reliabilities, Experiment 3 used a multiple-observation task, which allowed for estimating perceptual weights from multiple-cue stimuli. Using the same multiple-observation task, the integration of stereoscopic disparity, shading, and texture gradients was examined in Experiment 4. It turned out that less reliable cues were downweighted in the combined percept. Moreover, a specific influence of cue consistency was revealed. Shading and disparity seemed to be processed interactively while other cue combinations could be well described by additive integration rules. These results suggest that cue combination in visual depth perception is highly flexible and depends on single-cue properties as well as on interrelations among cues. The extension of the traditional cue combination model is defended in terms of the necessity for robust perception in ecologically valid environments and the current findings are discussed in the light of emerging computational theories and neuroscientific approaches.

> "It is plain that Distance is in its own Nature imperceptible, and yet it is perceived by Sight. It remains, therefore, that it be brought into View by means of some other Idea, that is it self immediately perceived in the Act of Vision."
>
> (Berkeley, 1732, SECT. XI.)

2. Introduction

2.1 Why sensory integration?

Numerous senses allow us to effectively interact with the world around us and we seem to effortlessly integrate these different sensory channels (e.g., vision, audition, touch, etc.) into a unified percept. To this aim, several sources of sensory information are simultaneously processed, analyzed and combined. Consider for example playing guitar. Besides viewing our fingers, we proprioceptively feel the position of our limbs. Tactile information helps us to decide on which string the fingers are positioned and which force is being exerted. Finally, we hear the sound when touching the strings. All these information are simultaneously collected and processed to contribute to our experience of the current situation. How is this sensory integration realized?

After conducting extensive research on perceptual systems in isolation, interdisciplinary research started to closely examine multisensory perception in the last decades (Stein & Meredith, 1993). Some basic characteristics of such integration processes seem to reflect generic principles of the nervous system that are comparable between intersensory integration (as described in the above mentioned example) and cue combination within one sensory system (Ernst & Bülthoff, 2004). The current study aimed at characterizing these principles within visual depth perception, which has attracted considerable attention throughout the last centuries (e.g., Berkeley, 1732) since depth is readily perceived although it has to be reconstructed from diverse visual cues that are subsequently integrated to obtain robust depth percepts.

2.1.1 Intersensory perceptual phenomena

Most of the early demonstrations of intersensory integration were described as perceptual illusions and the McGurk effect can certainly be regarded as belonging to the most impressive ones. McGurk and MacDonald (1976) asked children and adult observers to repeat what an audiovisually presented speaker said. Unbeknownst to the participants, visual (lip movement) and auditory information were discrepant. Interestingly, nearly all adults (98%) and 80% of the children reported hearing /da-da/ when viewing the utterance /ga-ga/ while hearing /ba-ba/. Thus, visual and auditory information were integrated into a combined percept that seems to represent a conglomerate of both sensory channels. This effect was found to be very robust and also occurred with different pairings of visual and auditory information (MacDonald & McGurk, 1978).

A second intersensory phenomenon that has received a considerable amount of attention is the redundant signal effect (cf., Kinchla, 1974). This effect characterizes an acceleration of response times to a signal that is presented simultaneously in two modalities as compared to each single modality. Originally, it was tried to explain this phenomenon by the so-called race model. This model assumes that several modalities produce independent activations, with the faster activation triggering the motor response (cf., Raab, 1962). However, several studies clearly demonstrated that multimodal responses were faster than predicted by the race model (e.g., Gondan, Lange, Rösler, & Röder, 2004; Miller, 1982, 1991). Thus, several modalities seem to be processed interactively instead of independently.

The above mentioned examples reveal two important aspects of sensory integration that will be discussed in detail below. First, the percept that emerges in these experiments usually shares features of the modalities that are combined. Thus, all currently available information about a stimulus or a scene seems to be effectively integrated. Second, the combined estimate is more reliable than each single cue. Thus, in using this combined information, the organism might deal more effectively with current environmental requirements.

2.1.2 Intrasensory integration in visual depth perception

Sensory integration does not only occur between modalities. Such processes are also highly relevant for combining information within one modality. With respect to such intramodal integration, the largest amount of research was conducted on visual depth perception. Humans are extraordinarily capable of estimating distances and dimensions of objects in their proximal space to allow for an efficient interaction with them. This capability is particularly surprising because it relies on a two-dimensional projection of the viewed scene on the retinas of both eyes. Consequently, a vivid depth perception can even be produced by flat pictures (see Figure 2.1; Koenderink, van Doorn, & Kappers, 1992). Thus, the question arises how a three-dimensional perception is generated from two-dimensional retinal images (cf., Todd, 2004).

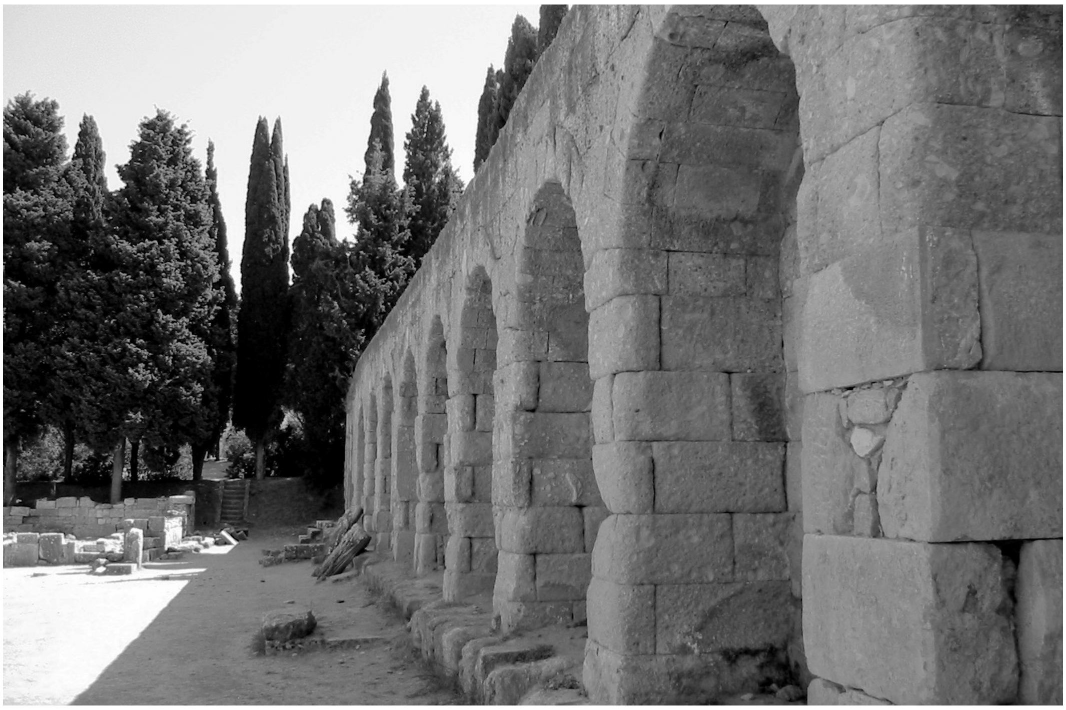


Figure 2.1. Illustration of depth perception in flat pictures. The photograph depicts an ancient wall of the ruins of the Askleipion on Kos, Greece. A vivid depth perception emerges mainly because of highly salient perspective cues and texture gradients. Additionally, the familiar size of various objects in the scene contributes to perceived depth.

Much research has focused on the so-called cue approach to explain this construction of spatial depth. To this aim, a number of information in the retinal images of both eyes were identified that correlate with depth in the world and it was reasoned that these cues can be extracted and interpreted by the observer to infer the scene’s three-dimensional structure (cf., Goldstein, 2007, chapter 8). These cues can be roughly partitioned into (1) pictorial depth cues, such as perspective, texture gradient and shading (see Figure 2.1), (2) stereoscopic (i.e., disparity) and ocular depth cues, and (3) motion-induced depth cues that either result from an observer movement (i.e., motion parallax) or an animation of the viewed scene (for a more comprehensive taxonomy see Palmer, 1999). Interestingly, an observer is typically unaware of using such information and moreover, the depth percept is highly monolithic. For example, when judging the distance of objects, we typically perceive a unitary spatial depth and not a fractionated pattern of multiple depth estimates. Thus, our sensory system apparently integrates several cues that encode spatial depth (cf., Jacobs, 2002a). The advantages of such integration are very similar to the intermodal case. The combined estimate utilizes all information that is present in a given scene to disambiguate single cues as well as to enhance the reliability of the combined percept (Ernst & Bülthoff, 2004).

2.2 Cue combination models

Despite the benefit that might result from sensory integration, early studies suggested that several signals are not integrated into a unified percept by combining the characteristics of single cues. Instead, one cue was found to completely determine the percept. In a study on visual-haptic shape perception, Rock and Victor (1964) used lenses to present stimuli with dramatically discrepant visual and haptic sizes (the visual width of the stimulus was reduced by approximately one half). Interestingly, most observers did not become aware of conflicting shape information when seeing and touching the object. But the combined estimate of both senses was strongly determined by the object's visual size and in fact, it was indistinguishable from the perceived shape in a solely visual control condition. Thus, cue vetoing instead of cue integration occurred. Comparable results were reported by Hay, Pick, and Ikeda (1965), and Singer and Day (1969). Moreover, cue vetoing was also observed for the intramodal case in the perception of three-dimensional layout from several visual cues (e.g., Bülthoff & Mallot, 1988; O'Brien & Johnston, 2000).

Later, these cases of cue vetoing were interpreted as resulting from either huge differences in the reliabilities of single cues in the respective experimental studies (e.g., Ernst & Banks, 2002, p. 432) or from using comparably large conflicts between individual cues that might prevent the perceptual system from integrating them into a combined estimate (e.g., Warren & Cleaves, 1971, p. 207; Gepshtein, Burge, Ernst, & Banks, 2005, p. 1020 f.). Thus, cue vetoing was though to represent only a special case of multisensory perception. Moreover, cue integration instead of vetoing was repeatedly reported for the perception of surface texture from visual and tactile cues (e.g., Jones & O'Neil, 1985; Lederman & Abbott, 1981; Lederman, Thorne, & Jones, 1986), for multisensory judgments of spatial location (e.g., Pick, Warren, & Hay, 1969), and for the combination of several visual depth cues into a unified percept (e.g., Johnston, Cumming, & Landy, 1994; Johnston, Cumming, & Parker, 1993; Young, Landy, & Maloney, 1993). Especially for the latter intramodal case, Landy and colleagues proposed a classification of combination rules that will be discussed below (Landy, Maloney, Johnston, & Young, 1995). These rules can be arranged on a continuum ranging from strong fusion to weak observer models (q.v., Parker, Cumming, Johnston, & Hurlbert, 1995). For the case of cue integration in visual depth perception, a hybrid model called modified weak fusion is outlined.

2.2.1 Strong fusion

Strong fusion models do not rely on the assumption of several separate modules that independently compute depth maps for the given retinal image. Thus, depth cues such as texture, shading, occlusion, stereoscopic disparity, etc. are thought to be "artificial" constructs which were needed to describe the results of specifically designed experiments. In the strong fusion model of Nakayama and Shimojo (1992), for example, an observer is thought to determine the most probable three-dimensional interpretation of a scene S given the current retinal image I. Thus, the scene perception corresponds to the maximum of the likelihood function $p(I \mid S)$. Within this framework, no modularization is required and consequently, no formal cue integration rule has to be established. Even when assuming several depth modules, they are supposed to interact closely in multiple ways to compute the overall depth map.

Unfortunately, strong fusion models are difficult to test empirically because they can be arbitrarily complex and the computational rules are not formalized. As a result, strong fusion models have not been tested directly, but instead, empirical data incongruent to a modular processing of cues have been thought to substantiate this form of sensory integration. For example, the combination of texture and motion cues in visual slant perception could not be modeled under the assumption of separate modules, thus, potentially providing evidence for a strong coupling of both cues (Rosas, Wichmann, & Wagemans, 2007).

2.2.2 Weak fusion

In weak fusion models, separate depth modules are assumed that independently compute depth maps from a given retinal image. According to the notion of Landy and colleagues (1995), these maps are subsequently averaged to obtain an overall depth map that corresponds to the observer's perception of the current scene. In contrast to strong fusion models, this framework can be empirically tested. For example, the modular structure allows for deriving predictions for the combined depth estimate on the basis of isolated depth cues.

The main problem of this integration scheme, however, is its implausibility. Depth cues qualitatively differ with respect to the information they provide and it is impossible to calculate quantitative depth maps on the basis of each single cue (Maloney & Landy, 1989). For example, motion parallax allows for absolute depth estimates, whereas texture gradients only provide relative depth information. Other cues, such as occlusion, only allow for judging the relative position of several objects in a given scene. Thus, each cue provides different and mostly incommensurate depth information and it seems unlikely or even impossible that separate depth maps are calculated and averaged on this basis. Moreover, a simple averaging would neglect that some cues might be more reliable than others in certain image regions. For example, texture can be highly reliable for regularly textured surface regions but not for regions containing little or noisy texture information (Rosas, Wichmann, & Wagemans, 2004). For any cue combination scheme, it seems plausible to take these variations of reliability into account instead of simply averaging depth maps. Thus, different flexible weights should be assigned to differentially reliable cues.

2.2.3 Modified weak fusion

The modified weak fusion model can be understood as a compromise between weak and strong observer models. Within this framework, a limited set of interactions between depth cues is included. These interactions serve to promote depth cues to allow for a calculation of commensurate depth maps. For example, the familiar size of an object provides distance information according to Emmert's law, which, in turn, can be used to promote stereoscopic disparity for computing absolute depth maps (O'Leary & Wallach, 1980; for other examples, see Landy et al., 1995, p. 391 ff.). Within the modified weak fusion scheme, it is assumed that separate, but commensurate depth maps are calculated for several cues by using such interactions. In a second step, these depth maps are integrated with respect to their reliability. This integration is assumed to be dynamic even within one scene. Thus, cues are locally integrated by using a flexible weighting scheme.

Several aspects of this modified weak fusion framework have been tested empirically. For example, cue promotion was observed when integrating stereo and motion depth cues (Johnston et al., 1994), and a computational simulation study further substantiated the superiority of the modified weak fusion scheme in this case (Fine & Jacobs, 1999). Moreover, a linear combination of depth cues was repeatedly demonstrated (e.g., Bruno & Cutting, 1988; Dosher, Sperling, & Wurst, 1986; Young et al., 1993). However, the most extensively examined aspect of this model concerns the reliability-sensitive integration of depth cues. These results, which are also highly relevant for the current study, will be discussed in detail in sections 2.4.1 and 2.4.2.

2.3 Bayesian models of cue integration

The above mentioned cue combination models as well as the veto rule can be integrated into a more global Bayesian combination scheme. Bayesian probability theory provides a normative framework for modeling how observers combine the information of multiple cues with prior knowledge about the world to make perceptual inferences (Ernst, 2006; Ghahramani, Wolpert, & Jordan, 1997; Knill & Richards, 1996; Mamassian, Landy, & Maloney, 2002). To illustrate this Bayesian integration scheme, assume that two cues encode the same property of an object, for example auditory and visual cues indicating an object's location (cf., Alais & Burr, 2004; Battaglia, Jacobs & Aslin, 2003). To estimate the location of the stimulus *S*, an observer should rely on the posterior probability function $p(S \mid s_1, s_2)$ with s_1 and s_2 representing the auditory and visual cues, respectively. The measurements associated with s_1 and s_2, depend on the stimulus *S* but are assumed to be corrupted by a certain amount of noise. Using Bayes' rule, the posterior probability function can be calculated by

$$p(S \mid s_1, s_2) = \frac{p(s_1, s_2 \mid S) \cdot p(S)}{p(s_1, s_2)} \tag{2.1}$$

with $p(\mathrm{s}_1,\ \mathrm{s}_2 \mid S)$ representing the likelihood function and $p(S)$ the prior. The denominator $p(s_1, s_2)$ does not depend on the given stimulus, thus, it is constant and equation 2.1 can be rewritten as

$$p(S \mid s_1, s_2) \propto p(s_1, s_2 \mid S) \cdot p(S) \tag{2.2}$$

This general rule that incorporates linear as well as nonlinear cue combination schemes (depending on the structure of likelihood and prior probability functions) can be used to develop predictive theories about the human sensory systems and these predictions can be tested psychophysically (Knill & Saunders, 2004). Figure 2.2 shows several examples of the flexibility of this Bayesian framework. In all three panels, the cues are conditionally independent (i.e., the sensory noise associated with each cue is independent). For both cues, likelihood functions are normally distributed with different means and σ_1^2 / σ_2^2 = 4. In panel A, the prior *p*(*S*) is flat so that the posterior equals the likelihood. In this case, the cues are not integrated and both estimates remain separately accessible. In panel B, cue 2 completely determines the posterior regardless of the sensory estimate of cue 1. This corresponds to the veto rule that is described above (see section 2.2).

Both cues are completely fused when the prior is proportional to Kronecker's delta function δ_{ij}

$$\delta_{\mathrm{ij}} = \begin{cases} 1, \text{ if } i = j \\ 0, \text{ if } i \neq j \end{cases} \tag{2.3}$$

Given that the noise of individual estimates are independent and Gaussian, the posterior equals the Maximum Likelihood Estimate (MLE) in this case. Thus, this integration scheme that is depicted in panel C of Figure 2.2, can be understood as an ideal observer model. It takes into account all cues that are available and minimizes the variance of the combined estimate (Cochran, 1937; Oruç, Maloney, & Landy, 2003, p. 2464 f.).

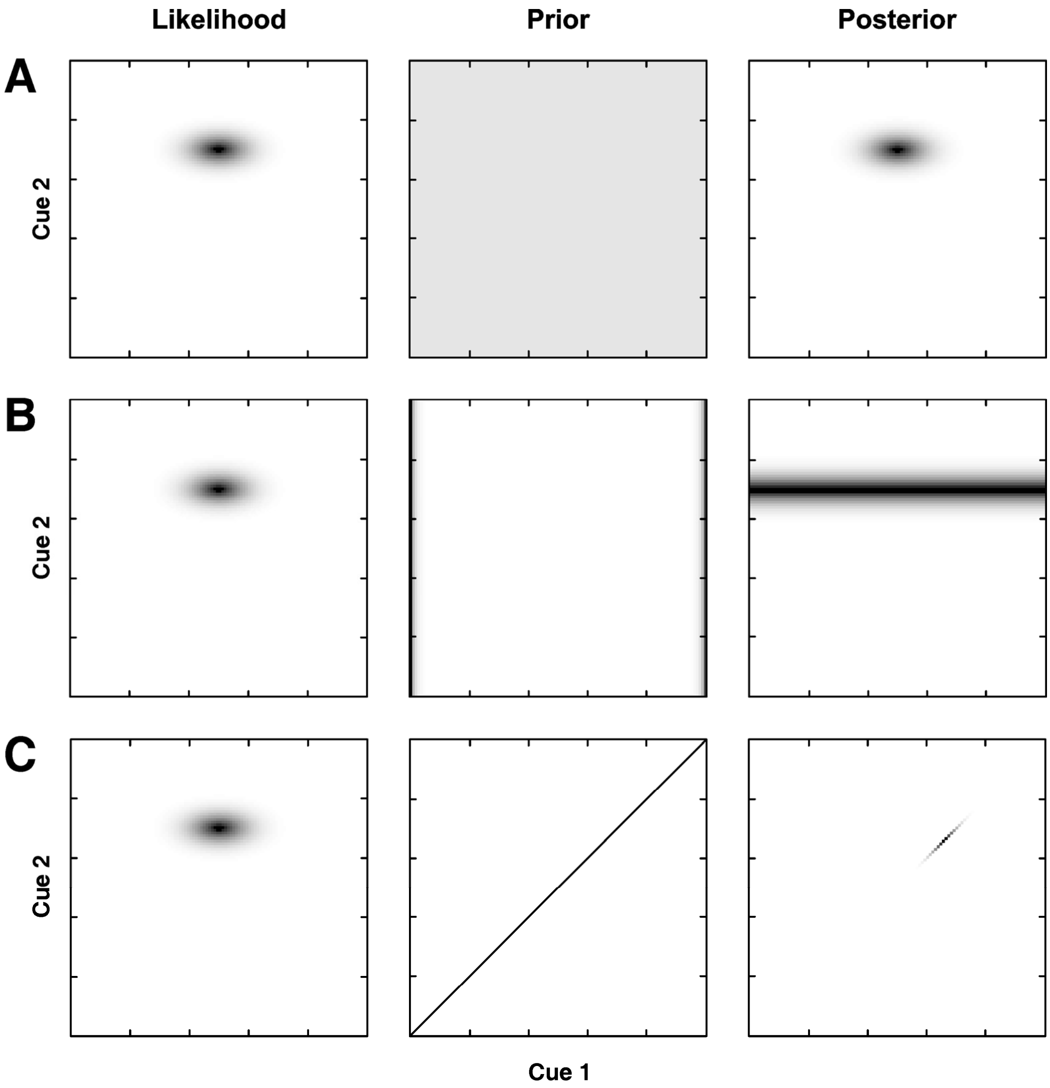


Figure 2.2. Illustrations of Bayesian cue combination schemes. The likelihood distributions that are shown in the left column depict the sensory information that is available to estimate a given stimulus property. The middle column shows the prior distributions that describe the mapping between sensory signals. The posterior, which is proportional to the product of likelihood and prior distributions, is depicted in the right column. Panel A shows results for a flat prior. Panel B depicts a veto rule with cue 2 completely determining the percept. In panel C, both cues are completely fused because the prior corresponds to a sharply tuned delta function (the identity line). In all panels, darker dots correspond to increasing probability.

2.4 The simple Maximum Likelihood Estimation (MLE) model

As outlined above, the MLE model is a special case of a Bayesian integration scheme where single cues are completely fused into an integrative estimate. In this case, the combined estimate $\hat{s}_\mathrm{c}$, which is the expected value of the posterior probability function, amounts to

$$\hat{s}_\mathrm{c} = \mathrm{E}\left[p(S \mid s_1, s_2)\right] = \frac{\sigma_2{}^2}{\sigma_1{}^2 + \sigma_2{}^2} \cdot \hat{s}_1 + \frac{\sigma_1{}^2}{\sigma_1{}^2 + \sigma_2{}^2} \cdot \hat{s}_2 \qquad (2.4)$$

with $\hat{s}_1$ and $\hat{s}_2$ being the estimates of both individual cues and $\sigma_1{}^2$ and $\sigma_2{}^2$ being the single-cue variances. The standard deviation of the combined estimate σ_c is

$$\sigma_\mathrm{c} = \sqrt{\mathrm{Var}[p(S \mid s_1, s_2)]} = \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1{}^2 + \sigma_2{}^2}} \qquad (2.5)$$

For the complete derivation of this equations see Appendix A. If the reliability r_i is defined to be the inverse variance of the sensory estimates i, equation 2.4 can be rewritten to

$$\hat{s}_\mathrm{c} = \frac{r_1}{r_1 + r_2} \cdot \hat{s}_1 + \frac{r_2}{r_1 + r_2} \cdot \hat{s}_2 \qquad (2.6)$$

and the reliability of the combined estimate r_c amounts to

$$r_\mathrm{c} = \frac{1}{\sigma_\mathrm{c}{}^2} = \frac{1}{\sigma_1{}^2} + \frac{1}{\sigma_2{}^2} = r_1 + r_2 \qquad (2.7)$$

For n mutually independent sensory estimates, these calculations can be generalized to

$$\hat{s}_\mathrm{c} = \sum_{i=1}^{n} \left(\frac{r_\mathrm{i}}{\sum_{j=1}^{n} r_\mathrm{j}} \cdot \hat{s}_\mathrm{i} \right) \qquad (2.8)$$

and

$$r_\mathrm{c} = \sum_{i=1}^{n} r_\mathrm{i} \qquad (2.9)$$

with

$$r_{\mathrm{i}} = \frac{1}{\sigma_{\mathrm{i}}^{2}} \qquad (2.10)$$

Two important features of this combination rule can be derived from these calculations. First, all sensory cues contribute to the combined estimate and they are weighted proportional to their inverse variances (i.e. their reliabilities). Thus, the model offers a straightforward understanding of differential cue weighting. Cues that cannot be evaluated very well by the observer would lead to very variable estimates of the respective physical property. Within the MLE approach, these cues would receive a comparably smaller weight because of their relatively large ambiguity. Cues that can be estimated with high precision, on the other hand, would lead to very stable estimates which would receive larger weights within the MLE framework (see Yuille & Bülthoff, 1996). Second, the reliability of the combined estimate is always larger than each single-cue reliability. This satisfies one major goal of sensory integration which is the minimization of uncertainty (Ernst & Bülthoff, 2004). The largest gain in the reliability of the combined estimate as compared to the single cues occurs when the variance of single-cue estimates and thus also their weights are equal. For the case of two cues, this ratio of maximal improvement is $\sqrt{2}/1$ (see Gepshtein et al., 2005, p. 1014 f.). These features of the MLE model are illustrated in Figure 2.3. The combined estimate is largely determined by Cue 2, which has the higher reliability (as indicated by a smaller variance of the corresponding Gaussian distribution). Moreover, the reliability of the integrated estimate is larger than both single-cue reliabilities.

The simple MLE model assumes single cues to be conditionally independent. However, this model can also be generalized to the more complex case of correlated cues. In this case, the maximum possible reliability of the combined estimate is lower as compared to the integration of independent cues[1]. However, even when using highly similar visual cues (linear perspective and

[1] Strictly speaking, this is only true for positively correlated cues that have positive single-cue weights that sum up to one (Oruç et al., 2003, p. 2462 ff.).

texture gradient) in a slant discrimination task, it turned out that cue correlations could be neglected for nearly half of the observers (Oruç et al., 2003). Potentially, the visual system is capable of de-correlating several signals which optimizes their integration with respect to the MLE framework (Barlow & Földiák, 1989).

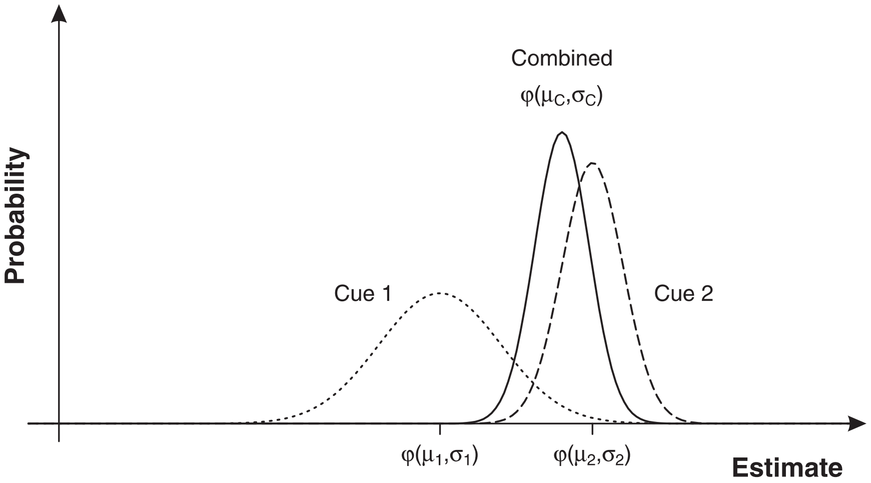


Figure 2.3. Illustration of cue combination according to the MLE model. Depicted are the probability density functions of two cues that differ with respect to expected value and standard deviation. The density function of the combined estimate is generated by normalizing the product of both single-cue functions (see Appendix A). All values correspond to the Bayesian integration scheme that is depicted in Figure 2.2, panel C.

2.4.1 Empirical evidence for a weighted linear integration model

The MLE integration scheme that is outlined above makes several predictions that can be tested empirically. In this section, empirical evidence that qualitatively supports the MLE model will be summarized. In the following section, recent studies that quantitatively probed the MLE integration scheme will be described (for a review, see Ernst & Bülthoff, 2004).

According to the MLE model, several sensory cues that encode the same physical property are integrated in a weighted linear manner. Moreover, single-cue weights are supposed to be directly proportional to their reliability. Thus, reducing the reliability of one cue should result in a reduced weighting in the combined percept. These qualitative predictions were supported by a large body of research.

For the intermodal case, van Beers and colleagues demonstrated that proprioceptive and visual location cues are integrated in a weighted linear manner (van Beers, Sittig, & Denier van der Gon, 1999). In their study, they took advantage of the asymmetry in the reliability of proprioceptive and visual location cues. Proprioceptive localization is generally more precise in the radial direction with respect to the shoulder than in the azimuthal direction. By contrast, visual localization is more precise in the azimuthal direction with respect to the cyclopean eye than in the radial direction (van Beers, Sittig, & Denier van der Gon, 1998). When integrating proprioceptive and visual position information according to the MLE rule, the two-dimensional probability distribution of the integrated estimate should resemble the multiplicative combination of both single-cue estimates (see Appendix A). This prediction was indeed verified by the respective study: The performance in the bimodal task could be well described by a weighted average of visual and proprioceptive cues taking into account the spatial distribution of single-cue reliabilities (cf., van Beers, Wolpert, & Haggard, 2002). Moreover, shifting the visual field by prism goggles led to a shift of the combined position estimate into the predicted direction.

Several studies were also carried out on intermodal event perception. These studies were mainly based on the finding of a surprisingly robust perceptual illusion that was described by Shams, Kamitani, and Shimojo (2000, 2002). These researchers presented a single visual flash accompanied by multiple beeps. Interestingly, their observers consistently reported having seen multiple flashes. Thus, visual and auditory information were integrated to a certain degree. This effect was robust to variation of many parameters (e.g., perceptual characteristics of the visual flash) but it disappeared when visual and auditory signals were presented with a temporal gap of at least 150 ms (Shams et al., 2002). Andersen, Tiippana, and Sams (2004) showed that this effect is basically bidirectional, thus auditory beeps can affect the number of perceived visual flashes and the number of flashes can influence the perceived number of beeps. Moreover, the integration was shown to be sensitive to the reliability of the auditory signal: Vision had a significant influence on audition only when the reliability of the auditory signal was reduced (q.v., Shams, Ma, & Beierholm, 2005). In a later reanalysis of these data, it was shown that both modalities seem to be integrated in a reliability dependent manner on the basis of continuous probability distributions, thus resembling important characteristics of the MLE rule (Andersen, Tiippana, & Sams, 2005).

These results also generalize to the integration of other modalities. Bresciani and colleagues (2005) demonstrated that auditory beeps and tactile taps are integrated in a comparable manner. The number of auditory beeps significantly influenced tactile tap perception at least when both signals were presented in temporal proximity. Bresciani, Dammeier, and Ernst (2006) focused on the direction of the integration and asked participants either to count visual flashes or simultaneously delivered tactile taps. The number of events in both channels could differ by one. It turned out that the perceived number of events depended on the number of events in the background modality that had to be ignored. Interestingly, the effect of touch on vision was stronger than vice versa and touch was also shown to be moderately more reliable when presented in isolation. Moreover, the variance of the observers' estimates in the multiple-cue condition was smaller than in the single-cue conditions (see Violentyev, Shimojo, & Shams, 2005, for comparable results). Thus, both signals seemed to be integrated in a reliability dependent manner as predicted by the MLE integration scheme. This was also confirmed by recent study focussing in the integration of auditory and tactile signals (Bresciani & Ernst, 2007). When lowering the reliability of the auditory cue (by reducing the intensity), its influence on the estimated number of events was reduced. Taken together, sensory estimates of different modalities in the perception of several distinguishable events seem to be integrated in a reliability dependent manner.

Comparable results were also reported for experiments that solely concentrated on the visual modality, thus examining intrasensory integration of multiple cues. Young and colleagues (1993) used cylinders as stimuli with texture and motion cues that could be varied independently of each other. Observers were asked to estimate the cylinder's depth and it turned out that both cues determined the percept in an additive linear manner. Moreover, when corrupting one cue by added noise, the weighted linear combination rule shifted in favour of the uncontaminated cue as predicted by the MLE model. Johnston et al. (1994) reported comparable results for a shape judgement task where texture and stereo cues were used to depict cylindrical and spherical objects. A linear integration rule assigning a lower weight to the texture cue predicted their empirical data fairly well. Furthermore, at larger viewing distances where the stereo cue becomes less reliable, the weight of the texture cue increased (q.v., Landy et al., 1995, p. 406 ff.). A similar additive integration rule that is sensitive to single-cue reliabilities was also found for stereoscopic slant perception (Backus, Banks, van Ee, & Crowell, 1999) and for the integration of texture cues (particularly perspective convergence) and stereo information in slant perception (Saunders & Backus, 2006).

Recently, Drewing and Ernst (2006) used a more active task to examine how different proprioceptive signals are integrated into a combined estimate. Observers touched artificial surfaces (as generated by a force feedback device) and had to estimate their degree of convexity. Two cues emerging from this active touch can be used to infer the shape of the touched object: When sliding a finger across a surface, the finger follows the geometry of the object (positional cue). At the same time, forces related to the local slope of the object act on the finger (force cue). In two experiments, it turned out that both cues affected perceived convexity in an additive manner. Furthermore, the relative weights that were assigned to both cues seem to implicate that the reliability of the position cue increased with curvature whereas the reliability of the force cue remained relatively stable across curvature changes. Altogether, the predictions of the MLE rule were qualitatively supported by a large body of research for the intermodal as well as the intramodal case.

2.4.2 Empirical support of the simple MLE model

The MLE model does also make quantitative predictions about cue weights and reliabilities that could be tested empirically. The technological process in recent years (e.g., virtual reality environments, force feedback devices) now allows for such experiments that make use of artificial stimuli whose dimensions could be manipulated precisely. To illustrate the general principle of such experiments, an influential study of Ernst and Banks (2002) should be described in more detail. These authors were interested in examining the integration of visual and haptic cues when judging the size of an object. Therefore, virtual horizontal bars were presented visually and could be grasped using a force feedback device. The visual cue could be corrupted by one of three levels of added noise to systematically reduce its reliability. Ernst and Banks (2002) measured the reliability of each single cue (derived from the discrimination threshold) and additionally constructed multiple-cue conditions by pairing differentially reliable visual cues with the haptic cue. In the multiple-cue conditions, the visual height of the bar differed slightly and unnoticeably from the simulated haptic height. By measuring the perceived height of the bar in this condition, cue weights could be determined and compared to the predictions of the MLE model that were derived from the single-cue conditions (see equation 2.6). Moreover, the reliability of size estimates in the multiple-cue task could be compared to the predicted reliability for this condition (see equation 2.7). It turned out that the reliability of the visual cue was reduced by added noise. This reduction should result in a downweighting of this cue in the multiple-cue condition and this prediction was fully confirmed by the experimental data with close quantitative correspondence of predicted and empirical weights. Moreover, the reliability of the combined estimate was larger than each single-cue estimate and it closely corresponded to the predicted values for all levels of visual noise. Thus, quantitative predictions of the MLE model were calculated on the basis of the single-cue conditions and these predictions were fully confirmed in the multiple-cue condition.

Using a very similar setup, it was shown that visual and haptic shape cues seem to be integrated according to the MLE model (Helbig & Ernst, 2007). Alais and Burr (2004; q.v., Banks, 2004) asked participants to estimate the spatial location of an event that provided visual and auditory cues and they also found that cue weights and the reliability of the combined stimulus could be precisely estimated from the single-cue conditions by using the MLE framework. Comparable qualitative results were also reported by Ghahramani et al. (1997), and Slutsky and Recanzone (2001). Moreover, Heron, Whitaker, and McGraw (2004) demonstrated that visual and auditory cues were integrated in a reliability dependent manner when observers had to judge the spatial location of moving objects.

Some studies also focused on the intramodal case and examined whether visual shape or depth cues are integrated in a similar manner (see section 2.1.2). Jacobs (1999), for example, separately measured discrimination thresholds for texture and motion cues that could be used to judge the depth of visual hemicylinders. These values were used to predict the perceived depth of multiple-cue stimuli on the basis of the MLE model. The multiple-cue stimuli consisted of texture and motion cues indicating a similar depth of the cylinder. It turned out that predicted and empirical depths were highly correlated, thus confirming an MLE-like integration rule. Unfortunately, the reliability of the combined estimate was not compared to the prediction of the MLE model but Figure 2 (Jacobs, 1999, p. 3626) indicates that it was smaller than single-cue reliabilities in some cases, thus questioning the general validity of this integration scheme (this point will be discussed in more detail in section 2.4.3).

Two recent studies examined the integration of texture and stereo cues in slant estimation with respect to the MLE model (Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003). Both made use of a known variation in texture reliabilities as a function of slant degree (Knill, 1998b; q.v., Rosas et al., 2004). Typically, they change by an order of magnitude from low to high slants, whereas the reliability of the stereo cue changes to a smaller degree. Thus the MLE model would predict a substantial upweighting of the texture cue as the slant increases.

This was confirmed by both studies with only small discrepancies from the quantitative predictions of the MLE model which may be due to the specific single-cue conditions that were chosen by Knill and Saunders (2003, p. 2555 f.). Moreover, Hillis et al. (2004) varied the distance between observer and slanted object to manipulate the reliability of the stereo cue which typically decreases as a function of distance (Johnston et al., 1993). Thus, a different pattern of cue weighting should be observed for nearby as compared to farther slants (Banks, Hooge, & Backus, 2001, p. 68 ff.) and this prediction was also verified by the experimental data. Both studies additionally compared the reliability of the combined stimulus to the predictions of the MLE model and reported a close quantitative correspondence between them.

Cue integration on a more basic level of visual processing was examined by Landy and Kojima (2001) who tried to predict the differential weighting of texture cues by referring to the MLE model. Observers had to accomplish a Vernier discrimination task, thus they had to judge the relative location of two edges. The edges were defined by spatial frequency and orientation of texture elements (Experiment 1) or line orientation and contrast (Experiment 2). All cues could be “blurred” to systematically manipulate their reliability. In both Experiments, the results of a multiple-cue condition were fully in line with an additive integration rule that is sensitive to single-cue reliabilities, thus providing evidence for an MLE-like integration scheme. However, the variance of the combined estimate was larger than the minimal variance of single cues for at least one of two observers. Such a result, which contradicts the MLE integration scheme, was also present in the data of Jacobs (1999) and the question arises whether the MLE model is really sufficient to fully explain the integration of cues across senses or even within one modality.

2.4.3 Empirical evidence conflicting with the simple MLE model

Although several studies provided empirical evidence supporting the simple MLE model, there are also a number of studies that conflict with this integration scheme. As already outlined in the previous section, some studies reported an MLE-like weighting of single cues, but the reliability of the integrated estimate fell short of the quantitative predictions of this integration model (e.g., Jacobs, 1999; Landy & Kojima, 2001). Opposite results revealing a larger reliability in the multiple-cue condition than predicted by the MLE model were also reported for the integration of proprioceptive and visual position information (van Beers, Sittig, & Denier van der Gon, 1996), for the integration of disparity and shading cues for surface interpolation (Vuong, Domini, & Caudek, 2006), and for shape detection and identification from the orientation and spatial frequency of Gabor patches (Meinhardt, Persike, Mesenholl, & Hagemann, 2006).

The MLE model did not only fail to predict the reliability of the combined estimate, there are also a number of studies reporting a substantial deviation of predicted and empirical weights. For example, Rosas and colleagues used a slant discrimination task with visual texture and haptic cues. They reported that cues tended to be weighted according to their reliability, but the observed weights in a multiple-cue condition substantially differed from the quantitative predictions of the MLE model (Rosas, Wagemans, Ernst, & Wichmann, 2005). Battaglia and colleagues (2003) used an audio-visual localization task which was roughly comparable to Alais and Burr (2004). The visual signal could be corrupted by one of five noise levels, thus, the weight assigned to the auditory cue should increase as a function of visual noise. Indeed, the authors found a reliability-sensitive weighting of single cues. However, the visual cue was weighted more strongly than predicted by the MLE model.

Comparable results conflicting with the MLE framework were also obtained in studies focussing on the intramodal integration of visual cues. Rosas et al. (2007) asked participants to judge the slant of surfaces using texture and motion cues. Although the observers seemed to use both cues for judging the slant’s surface, their response pattern could not be explained by the MLE model even when taking into account potential influences of correlated single-cue noises (Oruç et al., 2003). Even more obvious results were reported by O'Brien and Johnston (2000). With respect to the threshold data in a slant discrimination experiment, there was no consistent effect of adding motion to a given texture pattern. Almost no weight was attributed to the motion cue by the visual system. Thus, a veto rule, which is inconsistent with the MLE model, was employed. A control experiment verified that this response pattern was stable across different reliabilities of the texture cue. Furthermore, this cue vetoing was not due to extreme discrepancies between the reliability of texture and motion cues, respectively.

One aspect that seems crucial for the integration of multiple cues is their similarity. This aspect was demonstrated by Gepshtein and Banks (2003), for example. They used an experimental procedure that closely corresponds to studies focussing on the MLE framework (e.g., Ernst & Banks, 2002). Participants had to estimate the distance between two virtual surfaces either using vision or touch alone or by using both senses. The reliability of visual estimates was supposed to change as a function of object orientation with a better discrimination of surfaces being oriented parallel in contrast to perpendicular to the line of sight. Haptic precision should be unaffected by changes in object orientation, thus cue weights were supposed to change between both conditions when assuming an MLE-like integration scheme. Cue reliabilities were measured in separate single-cue conditions and these values were used to predict the perceived distance between the virtual surfaces and the reliability of the combined estimate in the visual-haptic condition using the MLE model. Slight conflicts between both senses were introduced in the multiple-cue condition to estimate the weights that were assigned to each cue (cf., Ernst & Banks, 2002). As predicted, the perceived distance was shifted towards the visual size when the surfaces were parallel to the line of sight and towards the haptic size when they were perpendicular, thus supporting a reliability-sensitive integration of both cues. Interestingly, the observed reliability in the visual-haptic condition was substantially smaller than predicted by the MLE-model. However, when confining the analysis to stimuli with small conflicts between both senses, the observed reliability was much closer to the predictions of the MLE-model. This discrepancy might be due to a disposition of the nervous system to integrate estimates only when they are assumed to stem from the same object. Thus, the consistency of several cues seems to be important for their integration. This assumption was further supported by a comparable study where the spatial offset between visual and haptic cues was systematically manipulated (Gepshtein et al., 2005). In this study, size discrimination of visual-haptic stimuli was most precise when visual and haptic signals were spatially coincident. Thus, the spatial separation between these signals determined how they were combined by the nervous system. This form of integration follows the proximity rule of Gestalt-Psychology (Wertheimer, 1923; see also Köhler, 1947, chapter V) and suggests that the MLE-model has to be extended by taking into account some form of similarity information of sensory estimates that seems to be necessary for initiating their integration.

The influence of cue similarity on their integration also seems to hold for the temporal domain. For example, the number of perceived flashes was only influenced by the (deviating) number of beeps when both signals were delivered in close temporal proximity (Shams et al., 2002). The same was true for the influence of auditory stimuli on the perceived number of tactile taps (Bresciani et al., 2005). Thus, close spatial and temporal proximity of several cues seems to be a necessary precondition for sensory integration within the MLE framework.

The above mentioned study of Gepshtein and Banks (2003), however, also suggests that this principle of similarity is much wider. Even when presenting several cues that are spatially and temporally coincident, they only seem to be integrated according to the MLE model when they are consistent to a certain degree. This was also convincingly demonstrated by Shams et al. (2005). In their study, observers viewed one to four flashes while hearing one to four beeps and they had to indicate the number of perceived flashes and beeps. It turned out that in conditions with small conflict between both channels, the number of events in one modality influenced the number of reported events in the other modality, thus sensory integration occurred. In conditions with large conflict, however, a smaller degree of integration and a larger amount of segregation was observed. Thus, the integration of auditory and visual cues clearly depended on the degree of conflict between both modalities. Highly comparable results were reported by Roach, Heron, and McGraw (2006) who examined audiovisual rate perception. They reported a gradual transition between partial cue integration and complete cue segregation with increasing inter-modal discrepancy.

Taken together, the simple MLE model failed to predict cue integration on an intermodal as well as an intramodal level in several studies. One factor that might have contributed to this finding is the lacking consideration of cue consistency as an important influence on cue combination strategies. An extended Bayesian model that includes cue interactions will be discussed below (see section 2.5).

2.4.4 Top-down influences on sensory integration

The simple MLE model described above is entirely bottom-up. That is, characteristics of the cues that are integrated fully determine the probability distribution of the combined percept. There are, however, some demonstrations of top-down effects on cue integration that question whether an entirely bottom-up model can fully account for the experimental data. Most of these experiments concentrated on the effect of former experience on the process of integration. Atkins, Fiser, and Jacobs (2001), for example, demonstrated that those visual cues that were experienced to be consistent with a haptic cue to spatial depth received a larger weight during cue integration than visual cues dissenting from the haptic estimate. This reweighting was shown to be context-dependent and occurred even in a naturalistic task without trialwise feedback. Atkins et al. (2001) concluded that “haptic percepts provide a standard against which the relative reliabilities of visual cues can be evaluated” (p. 459). In a second study, Atkins, Jacobs, and Knill (2003) showed that inconsistencies between stereo and haptic cues to spatial depth led to a recalibration of the stereo information, that is, depth-from-stereo estimates were aligned to match depth-from-haptic judgements after performing a training session with inconsistent depth cues. This recalibration was also shown to be context-dependent. Ernst, Banks, and Bülthoff (2000) demonstrated that consistencies between haptic feedback and visual stereo or texture information in a learning phase of a slant discrimination study increased the weight of the respective visual cue in a subsequent test phase. Recently, Ernst (2007) showed that introducing arbitrary correlations between vision and touch are capable of influencing their integration. He used virtual objects that consisted of two cues that are usually unrelated in the world (luminance and stiffness) and measured discrimination accuracy of bimodal stimuli before and after an extensive training period. In this training, both cues were either positively or negatively correlated depending on the group the observer was randomly assigned to. It turned out that discrimination accuracy in the post-test was exclusively enhanced for objects congruent to the training condition. Thus, the observers effectively learned to integrate the formerly unrelated signals (q.v., Jäkel & Ernst, 2003).

An experience-dependent reweighting was also shown for the intramodal case. Jacobs and Fine (1999) asked participants to judge the depth of virtual cylinders that provided texture and motion cues. Unbeknownst to the observers, one cue was fixed to depict a circular cross-section in a training condition, whereas the other varied from trial to trial. Thus, only one cue could be successfully used to infer the cylinder's depth. It turned out that this relevant cue was upweighted after prolonged training. Two other experiments showed that different rules of cue relevance could be learned for short and tall cylinders but not for cylinders depicted from different viewpoints. Thus, although context sensitive cue weighting occurred, some interrelations seemed to be more easily learned than others (cf., Haijiang, Saunders, Stone, & Backus, 2006).

Task requirements were also repeatedly shown to influence the process of cue integration in addition to specific stimulus features. Säfström and Edin (2004) asked participants to grasp a small object. Unbeknownst to them, seen and grasped object were never physically identical, thus, conflicts between both cues could be generated. After some initial trials, the haptic size of the object was either reduced or increased by 15 mm and the maximum grip aperture when reaching for the object was measured using an infrared tracking system. It turned out that changes in grip aperture closely followed the experimental manipulation of the object's haptic size. This change was more pronounced when the size increased, thus, haptic information was weighted more heavily when it was functionally more important for successfully grasping the object. Interestingly, observers who became aware of the conflicts between visual and haptic cues showed exactly the same response pattern as the group of observers that did not notice cue conflicts.

An even more obvious demonstration of task-dependent cue weighting was recently reported by Knill (2005). The relative influence of monocular (texture) and binocular cues in slant perception were measured in a between-subject design. One group of participants had to precisely place a virtual object on a slanted surface. Two other groups were instructed to adjust a probe to be perpendicular to the same surface by using either visual or proprioceptive information. Most importantly, both latter tasks were mainly perceptual whereas the object placement required some form of visuomotor coordination and movement dependent adjustment. It turned out that binocular cues contributed much more to the observer's performance in the object placement task than they did in either of the perceptual tasks. Thus, a qualitatively different cue-weighting strategy was used for motor control as compared to the computation of perceptual representation. This task-dependent weighting might be related to temporal differences in the processing of monocular and binocular cues (Greenwald, Knill, & Saunders, 2005).

In one of few studies using three different cues that could be separately manipulated to indicate a virtual surface, Tittle, Norman, Perotti, and Phillips (1998) showed that visual cues seem to be differentially integrated for the perception of scale-dependent (curvedness) and scale-independent (shape) surface features. Quadratic surfaces that varied in shape and curvedness served as stimuli. The surfaces were defined by shading, texture, and stereo cues which could be varied independently of each other. Observers had to adjust a stimulus with consistent cues to match a cue conflict stimulus. It turned out that both, judged shape and judged curvedness, could well be described by a weighted linear integration of all three cues. However, binocular disparity contributed significantly more to judged shape than to judged curvedness, whereas both monocular cues contributed more to judged curvedness. Thus, the integration of disparity, texture, and shading significantly varied according to whether scale-independent of scale-dependent information about the surface structure was extracted from the stimuli.

To sum up, former experience and task relevance clearly affect cue integration on the intermodal as well as the intramodal level. These top-down influences are not considered in the simple MLE framework but they can be integrated into a more general Bayesian integration scheme as will be shown below.

2.5 Bayesian integration of coupled cues

As discussed above, despite a large body of supporting evidence, the simple MLE model failed to predict the data of several studies on intersensory as well as intrasensory integration. However, this model can be regarded as representing only a special case of a more general Bayesian integration scheme and several researchers already tried to fully use the flexible Bayesian framework to incorporate incongruent results. This can most easily be implemented by modelling the influence of certain priors on the posterior probability distribution (see section 2.3). Battaglia et al. (2003), for example, found an enhanced weighting of visual cues in an audio-visual localization task that did not correspond to the predictions of the MLE model. However, by introducing a prior probability distribution that leads the model to make greater use of visual information, the empirical data could be predicted fairly well. This prior may be interpreted as some sort of prior preference for using visual signals in target localization tasks even when additionally available auditory cues may render the combined estimate more reliable (cf., visual capture, Rock & Victor, 1964).

The validity of the full Bayesian model in the integration of ambiguous shape cues was recently demonstrated by Adams and Mamassian (2004). Observers in this study were instructed to judge the depth of vertical ridges that were depicted by texture and disparity cues which could be discrepant to each other. When using only texture cues, the task was ambiguous because concave and convex stimuli generated the same pattern of texture gradients. By using the additional stereo cue, however, this ambiguity might be resolved. In this study, the observers' response pattern could be fully described by using a Bayesian model consisting of likelihood distributions for texture and disparity indicated depth, and separate prior distributions for convexity and flatness. These latter features represent a general preference of perceiving objects as being convex rather than concave (Liu & Todd, 2004), and a tendency for underestimating depth in this limited cue situation (Young et al., 1993). These examples of intersensory and intramodality integration illustrate that priors might have a considerable effect on cue integration that could be modeled with reference to the Bayesian framework.

In section 2.4.3, it was already speculated that some results that were inconsistent with the simple MLE model might be related to its lacking consideration of cue consistency as an important influence on cue combination strategies. In reality, an indiscriminate integration of sensory signals regardless of their relation and significance is certainly inappropriate. Consider, for example, hearing a telephone ringing behind you while watching television. In this case, it would be unfavorable to integrate the bell into the audiovisual input from the television and in fact, such highly conflicting cues (with respect to their spatial location, in this example) are rarely integrated into a combined estimate (Gepshtein & Banks, 2003; Gepshtein et al., 2005; Knill, 2007). Such consistency dependent integration is particularly important for enabling a robust integration of multiple cues (Landy et al., 1995, p. 394 f.; Maloney & Landy, 1989, p. 1160 f.). Roach et al. (2006) as well as Shams et al. (2005) demonstrated that audiovisual cues are indeed robustly integrated. Increasing the conflict between both channels led to an upweighting of the task-relevant modality. Moreover, a high degree of sensory fusion only occurred for stimuli with small cue conflicts. Such behavior of the sensory system can well be modeled within the Bayesian framework by introducing a coupling prior. However, this sort of coupling prior only defines the degree of sensory fusion as a function of cue conflict and does not affect relative cue weighting (cf., Bresciani et al., 2006; Ernst, 2006). Thus, highly reliable cues are assumed to have a large impact on the combined estimate regardless of whether they are consistent or discrepant to other cues at hand. With respect to robust fusion, however, it would make sense to downweight even highly reliable cues when they deviate from the majority of other sensory information about a specific stimulus property.

How can a Bayesian prior be constructed to allow for such consistency dependent cue weighting? For the case of three cues with expected values $\hat{s}_1$, $\hat{s}_2$, and $\hat{s}_3$ this might be achieved by using a three-dimensional Gaussian density function that is centered to the means of all cue pairings and has a different spread in all three spatial directions (σ_{12} in the direction of Cue 3; σ_{13} and σ_{23} in the direction of Cue 2 and 1, respectively). Thus, this prior is related to single-cue discrepancies from cue pairings and can be mathematically described by

$$\begin{aligned} p(S \mid s_1, s_2, s_3) = k \cdot e^{-\frac{1}{2}\left(\frac{s_3-(\hat{s}_1+\hat{s}_2)/2}{\sigma_{12}}\right)^2} \cdot \\ e^{-\frac{1}{2}\left(\frac{s_2-(\hat{s}_1+\hat{s}_3)/2}{\sigma_{13}}\right)^2} \cdot \\ e^{-\frac{1}{2}\left(\frac{s_1-(\hat{s}_2+\hat{s}_3)/2}{\sigma_{23}}\right)^2} \end{aligned} \tag{2.11}$$

k is a normalization constant that ensures that the integral of this prior probability density function is equal to 1. The standard deviations σ_{12}, σ_{13}, and σ_{23} define the degree of cue coupling with respect to their influence on relative cue weighting. A small value of σ_{12}, for example, implies that the combined estimate is strongly influenced by this cue pairing even when Cue 3 has a large reliability. This corresponds to a downweighting of Cue 3 which is illustrated in Figure 2.4. As can be seen from the likelihood distribution on the left side, Cue 3 has a high reliability whereas both other cues indicate consistent values (their two-dimensional distribution is centered on the diagonal) which are less reliable, however. According to the prior distribution in the middle column of Figure 2.4, Cue 1 and 2 are highly coupled (σ_{12} is small). Thus, the combined estimate is drawn into the direction of these highly coupled cues (see posterior distribution in the right column).

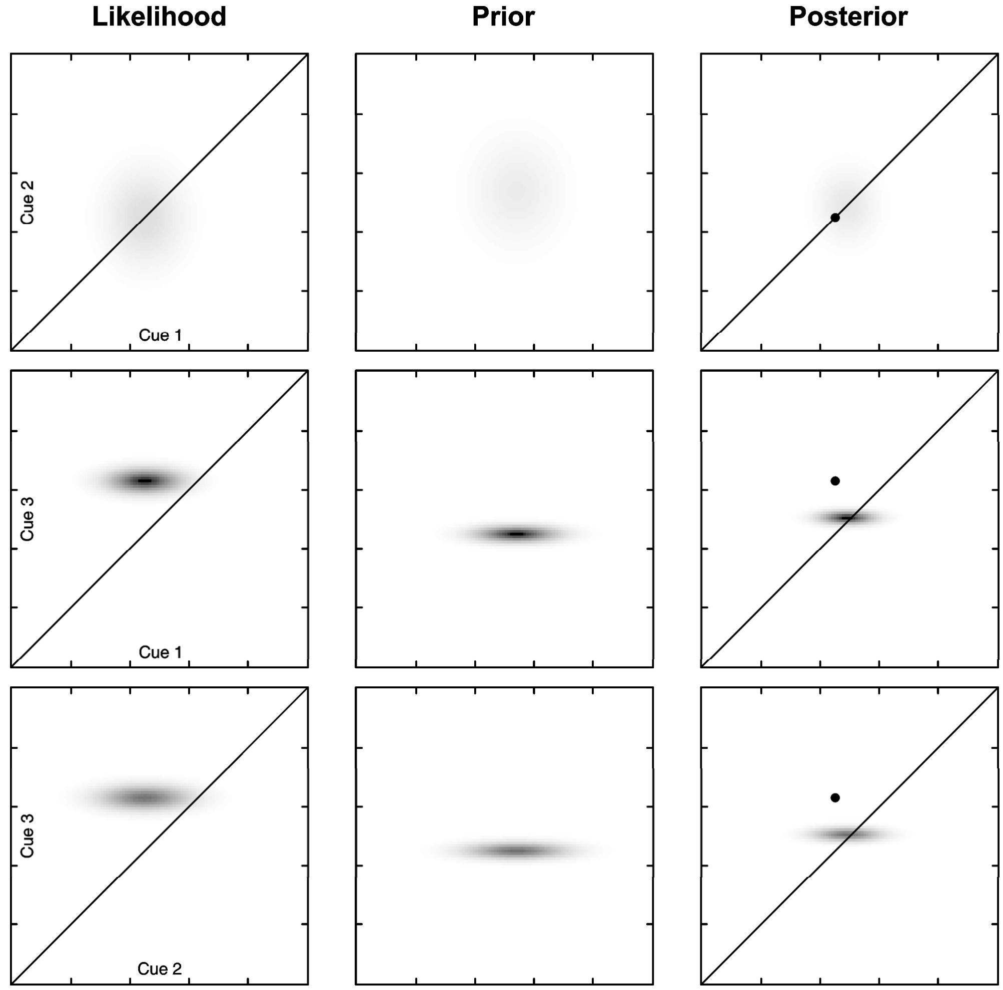


Figure 2.4. Bayesian cue combination of three differentially coupled cues. Depicted are two-dimensional projections of the probability distributions of each cue pairing. The diagonal represents points of cue similarity. The black dots in the panels of the posterior probability distribution depict the maximum probability of each cue likelihood to allow for an estimation of relative cue influence. In all panels, darker dots correspond to increasing probability.

On the basis of the MLE model, the combined estimate would be largely determined by the most reliable Cue 3 in this situation because cue consistencies are not taken into account (see Figure 2.5, left panel). According to the extended Bayesian model using a coupling prior as defined above, cues are not completely fused. Such incomplete integration fits to empirical data demonstrating that single-cue estimates remain accessible in a multiple-cue situation (e.g., Andersen et al., 2004; Bresciani et al., 2006; Hogervorst & Brenner, 2004; Shams et al., 2005). Furthermore, it was repeatedly shown that top-down effects have a considerable influence on cue integration strategies (see section 2.4.4). Especially task-dependent cue integration (e.g., Bradshaw, Parton, & Glennerster, 2000; Knill, 2005) might be interpreted as being based on an incompletely fused posterior probability distribution. Recently, Marian (2007) showed that attention influences cue integration strategies. In a depth matching task, the observer's attention was directed to specific stimulus attributes and it was shown that these attended cues were upweighted after cue consistencies were taken into account. Thus, it was demonstrated that cues seem to be separately accessible in the posterior distribution which supports the established integration scheme (cf., Hogervorst & Brenner, 2004).

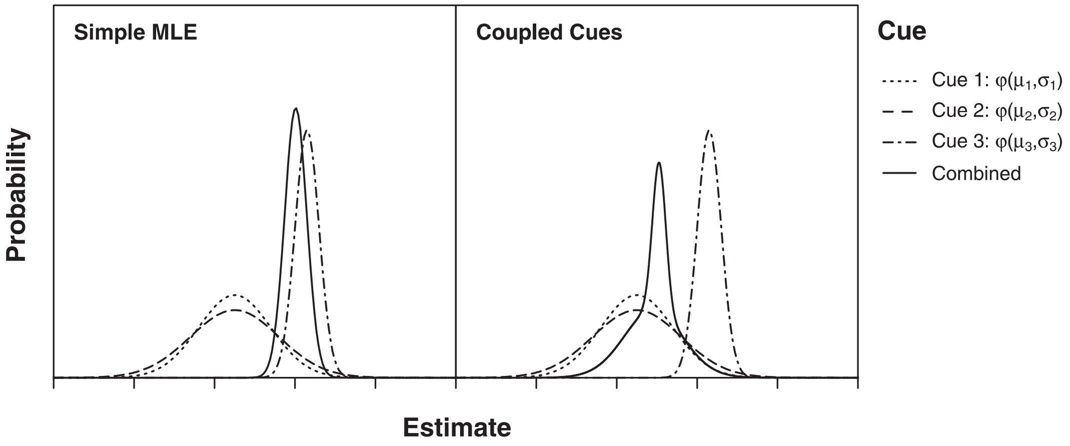


Figure 2.5. Comparison of cue integration according to the simple MLE framework (left panel) and an enhanced model that takes into account cue consistency (right panel). The Gaussian distributions of each single cue correspond to the values that are depicted in Figure 2.4.

But how do observers judge a given stimulus property (e.g., its depth or its spatial location) when separate single-cues still indicate different values in the posterior distribution? In the simplest case, the observer randomly selects one of the Gaussian distributions that result from a multiplication of likelihood and coupling prior and answers according to this probability distribution. In the case of three cues, her response pattern for a large number of trials would correspond to a normalized summation of the three Gaussian distributions lying in the main axes of the posterior centered to its global maximum. This case is depicted by the solid line in the right panel of Figure 2.5 using the same values as in Figure 2.4. As can be seen in direct comparison to the MLE integration scheme, the combined estimate is more strongly influenced by the consistent Cues 1 and 2 instead of the highly reliable Cue 3.

The variance of the combined distribution in the Bayesian model that takes into account cue consistencies can be smaller or larger than the corresponding value of the MLE model. It depends on the similarity of single cues as well as the strength of the coupling prior. Such flexibility allows for incorporating conflicting empirical results. Several studies reported smaller as well as larger reliabilities than predicted by the MLE framework (e.g., van Beers et al., 1996; Landy & Kojima, 2001; Vuong et al., 2006). These discrepancies might be related to a varying degree of cue coupling that was not taken into account.

Even the results of studies on the influence of former experience on cue integration strategies might be embedded within this Bayesian integration scheme by assuming that the coupling prior flexibly changes with experience (e.g., Adams, Graf, & Ernst, 2004; Atkins et al., 2001; Ernst et al., 2000). Thus, when two cues are repeatedly perceived as being correlated, their degree of coupling might increase which leads to a larger influence of these coupled cues on the combined estimate.

Taken together, the extended Bayesian model offers explanations for several studies that reported inconsistent results with respect to the simple MLE framework. The main disadvantage of this model, however, concerns its higher complexity. On the one hand, the MLE has no degrees of freedom. That is, all parameters are uniquely specified by their respective single-cue reliabilities. The extended Bayesian model on the other hand has three degrees of freedom in the above mentioned stimulus configuration. Thus, when deliberating about the adequacy of these models to characterize human perception in multi-cue situations, this differential complexity has to be taken into account.

2.6 Open research questions

Although consistencies between cues seem to affect their integration (Jacobs, 2002b), this aspect was not systematically taken into account until now in terms of a formal cue integration model. Moreover, it is still unclear whether such consistencies are detected and processed by default on a trial by trial basis or whether consistency based reweighting strategies have to be acquired through extensive training. Most available studies in this domain used long training periods to provide evidence for a consistency dependent cue weighting (e.g., nearly 700 training trials in Experiment 1 of Atkins et al., 2001). Thus, although these studies demonstrated that the perceptual system is highly flexible and adjusts cue weights to take into account consistencies, they reveal little insight into the process of robust integration (Maloney & Landy, 1989). To allow for such robust integration of several cues, the perceptual system has to quickly detect and process cue discrepancies when judging specific stimulus attributes. Triesch, Ballard, and Jacobs (2002) showed that observers can quickly detect varying cue reliability to adjust their weighting accordingly but it remains unclear whether this is also true for different patterns of cue consistency. Thus, the question emerges whether cue consistencies instantly affect sensory integration.

Moreover, it has yet to be demonstrated that correlated cues within one sensory system are similarly capable of influencing a combined estimate than are consistencies between haptic and visual cues that were examined in previous studies (Adams et al., 2004; Atkins et al., 2001, 2003; Ernst et al., 2000). It might be speculated that haptic cues are somewhat special because they possibly provide a “ground truth” against which visual cues are recalibrated. For example, children were supposed to learn how to interpret visual scenes by evaluating their motor interactions with the physical objects (see Piaget & Inhelder, 1956). Thus, the perceptual system might be specifically prepared to utilize visual-haptic interrelations and it is unclear whether these principles generalize to an intramodal level.

Cue consistencies were primarily examined with respect to the weight that different cues receive in a combined stimulus (e.g., Jacobs & Fine, 1999). However, cue interactions may also alter the reliability of the integrated estimate. On the one hand, the reliability might increase because the interrelation of several cues is taken into account in addition to single-cue information. On the other hand, single-cue estimates that indicate consistent properties of the stimulus at hand are redundant and thus may be associated with a reduction of estimation accuracy. It has still to be examined whether the reliability of the combined stimulus is reduced or enhanced when cue consistencies occur.

2.7 Outline of the present studies

The above mentioned research questions are directly linked to the extended Bayesian framework that includes a coupling prior (see section 2.5). This model might be a good candidate for explaining intramodal robust fusion but it has not yet been tested empirically. For the following experiments, visual displays of hemi-cylinders with an elliptical cross section were constructed whose depth was defined by at least three visual cues. In Experiment 1 and 2, the following depth cues were used: 1) a shading gradient was constructed by illuminating the cylinder with a laminar light source. 2) Circular texture elements were mapped onto the cylinder’s surface, thus producing different degrees of texture compression. 3) The texture elements moved along the cylinder’s surface, thus, their resulting velocities could be used to estimate the cylinder’s depth. The depth indicated by all three cues could be varied independently from each other. Consequently, this setup allowed for creating arbitrary consistencies between pairs of these cues. Following the procedure of former studies on cue integration within the MLE framework (e.g., Ernst & Banks, 2002), single-cue reliabilities were measured and these values were used to predict cue weights and the reliability of the combined estimate. These values were subsequently used to test the predictions of the simple MLE model against the extended Bayesian framework that takes cue consistencies into account.

In Experiment 3, a psychophysical procedure aiming to derive single-cue weights from a multiple-cue condition (Berg, 1989) was tested within a visual depth estimation task. This method was implemented to circumvent a major problem that was inherent in former studies on intramodal cue integration (e.g., Hillis et al., 2004; Jacobs, 1999; Knill & Saunders, 2003). These studies relied on the critical assumption that reliabilities of single cues might be measured in isolation and that these values are transferable to a multiple-cue condition (see Ernst & Bülthoff, 2004, p. 165 f.; Knill & Saunders, 2003, p. 3 f.). Experiment 3 examined whether single-cue weights can be directly determined from a multiple-cue condition without using specifically designed single-cue tasks. In addition to the three depth cues that were used in Experiment 1 and 2, stereoscopic disparity was included in the displays.

Finally, Experiment 4 used the psychophysical procedure that was successfully implemented in Experiment 3 to test whether single-cue reliability and multiple-cue consistency affect the integration of visual depth cues. In contrast to the former experiments, the reliabilities of single cues could be manipulated systematically in this study. Because of a potential confound of texture and motions cues (see Jacobs, 1999, p. 4064), static stimuli were used in this experiment. Thus, shading, texture, and stereoscopic disparity were used as depth cues. This setup additionally allowed for testing the generalizability of the conclusions that were derived from Experiment 1 and 2. In a first step, the data of Experiment 4 was qualitatively evaluated with respect to the adequacy of a simple linear integration model that is sensitive to cue reliabilities. Moreover, an MLE model was constructed on the basis of the multiple observation task and the empirically derived response pattern was quantitatively tested against the predictions of this integration scheme.

3. Experiment 1

3.1 Introduction

Most studies on visual cue integration have concentrated on the combination of two cues, for example, texture and motion (Jacobs, 1999; Young et al., 1993), stereopsis and kinetic depth (Johnston et al., 1994), or texture and stereoscopic disparity (Hillis et al., 2004; Johnston et al., 1993; Knill & Saunders, 2003). These studies demonstrated that single-cue reliabilities influenced the weighting of both cues when integrating them into a combined percept. Moreover, most of these data could well be described by the simple MLE integration scheme (see section 2.4.2). Unfortunately however, these studies provide little insight into the processing of cue interactions with respect to robust estimation of visual depth in complex and variable scenes. Already in 1995, Landy and colleagues proposed that "a stronger test of robustness would require stimuli containing three or more strong depth cues, one of which signals a depth discrepant with the depths signaled by the others." (Landy et al., 1995, p. 409). This proposal was adopted in the current study. Displays of visual hemicylinders were constructed that allowed for an orthogonal variation of shading, texture, and motion depth cues. By realizing arbitrary consistencies between these cues, the influence of cue interactions on the combined percept and thereby robustness could be evaluated. The current study closely followed the procedure of psychophysical studies on the MLE model (e.g. Ernst & Banks, 2002) to allow for a comparison of this integration scheme to an extended Bayesian framework that takes cue interactions into account (see section 2.5). In the following sections, the latter Bayesian integration scheme will be called consistency augmented MLE (CMLE) model because it enhances the MLE framework by explicitly modeling the influence of cue interactions on the process of sensory integration.

3.2 Method

3.2.1 Stimuli and apparatus

The stimuli were hemi-cylinders with an elliptical cross section rendered on a two-dimensional video display using parallel projection. One semiaxis of the elliptical cross section, which was always parallel to the image plane, was held constant (300 pixels[2]) and represented the cylinder's width. The other semiaxis, which was parallel to the observer's line of sight, corresponded to the cylinder's depth and was varied across trials. The cylinder's height was 600 pixels and equaled the vertical screen resolution.

The depth of the cylinder was defined by three visual cues: 1) The uniformly white surface of the cylinder with Lambertian reflectance properties was illuminated by a laminar light source from the direction of the observer, which produced a shading gradient. This gradient was normalized, thus, the part of the cylinder's surface that was nearest to the observer was always bright white, and the edge of the cylinder that was farthest was as black as the background of the screen. The gradient's shape could be used to extract depth information from the display (Mingolla & Todd, 1986; Pentland, 1989). 2) A homogeneous texture consisting of black circles was mapped onto the cylinder's surface. In a first step, between 80 and 120 black circles were randomly placed on a two-dimensional sheet that was sized 1000 × 600 pixels. The radius of each circle was randomly sampled from a uniform distribution ranging from 1 to 15 pixels. The texture generation algorithm prevented overlap among circles and changed their placement if necessary, such that the distance between the boundaries of any two circles was at least 20 pixels. The final texture was mapped onto the surface of the shaded cylinder. Texture elements were always coloured uniformly black and the shading gradient was visible in the space between them (see Figure 3.1). Different depths of the cylinder led to changes in size and shape of the texture elements as

[2] Given monitor size and screen resolution, one pixel amounted to approximately 0.455 mm × 0.455 mm. Thus, the width of the displayed cylinder on the screen was 136.5 mm.

well as in their density in certain parts of the display. This information can be used by the human visual system as a cue to an object's depth (Blake, Bülthoff, & Sheinberg, 1993; Buckley, Frisby, & Blake, 1996; Stevens, 1981; Todd & Akerstrom, 1987). 3) All texture elements could move horizontally along the surface of the cylinder with a constant velocity either in clockwise or counterclockwise direction. This was achieved by computing the position of the centre of each texture element as a function of time when travelling on a track defined by the elliptical cross section of the visible hemi-cylinder. The velocity was drawn from a uniform distribution that ranged from 132 to 162 pixels per second. It was always constant for each texture element within one trial, but it varied between trials to prevent that depth judgements could be based on absolute speed and maximum displacement of texture elements within the scene (Johnston et al., 1994). It is important to note that the cylinder itself did not rotate, only the texture elements moved along its surface. The same kind of motion information could be produced by wrapping a paper with black circles around a pillar and pulling it back and forth. Relative velocity and acceleration of texture elements in the centre and the edge of the cylinder are an indicator of the cylinder's depth when displaying this stimulus on a two-dimensional screen (Perotti, Todd, Lappin, & Phillips, 1998). Relatively flat cylinders, for example, lead to small differences between the velocities in the centre and at the edge. In deep cylinders, on the other hand, the texture elements seem to move slower at the edge compared to the cylinder's centre.

The algorithms that were used to generate the visual cues to the cylinder's depth worked independently from each other. Thus, the depth cues could be varied orthogonally. This technique allowed for an examination of the impact of single cues on the perceived depth of the cylinder.

Stimuli were computed in real time on a Pentium 4 2.4 GHz computer with a Radeon 7000 graphics chip using OpenGL via the Simple DirectMedia Layer cross-platform multimedia library (www.libsdl.org). They were displayed on a 19" monitor (Samtron 96P) with a resolution of 800 × 600 pixels and a color depth of 32 bits. The video refresh rate was set to 60 Hz, and it was made sure that the computer was fast enough to provide a frame rate for the animation of at least 60 Hz, too. The cylinder's width was set to 300 pixels and its height was 600 pixels, thus, the cylinder's bottom and top were never visible to the observer. The background luminance was 0.12 cd/m^2 when a stimulus was displayed. The luminance in the centre of the shading gradient, that is at its brightest point, was 110.40 cd/m^2. The computer was placed in a dimly lit room, and stimuli were viewed binocularly[3] from a distance of 1 m. The displayed image of the cylinder subtended 7.8° of visual angle in the horizontal and 15.6° in the vertical dimension.

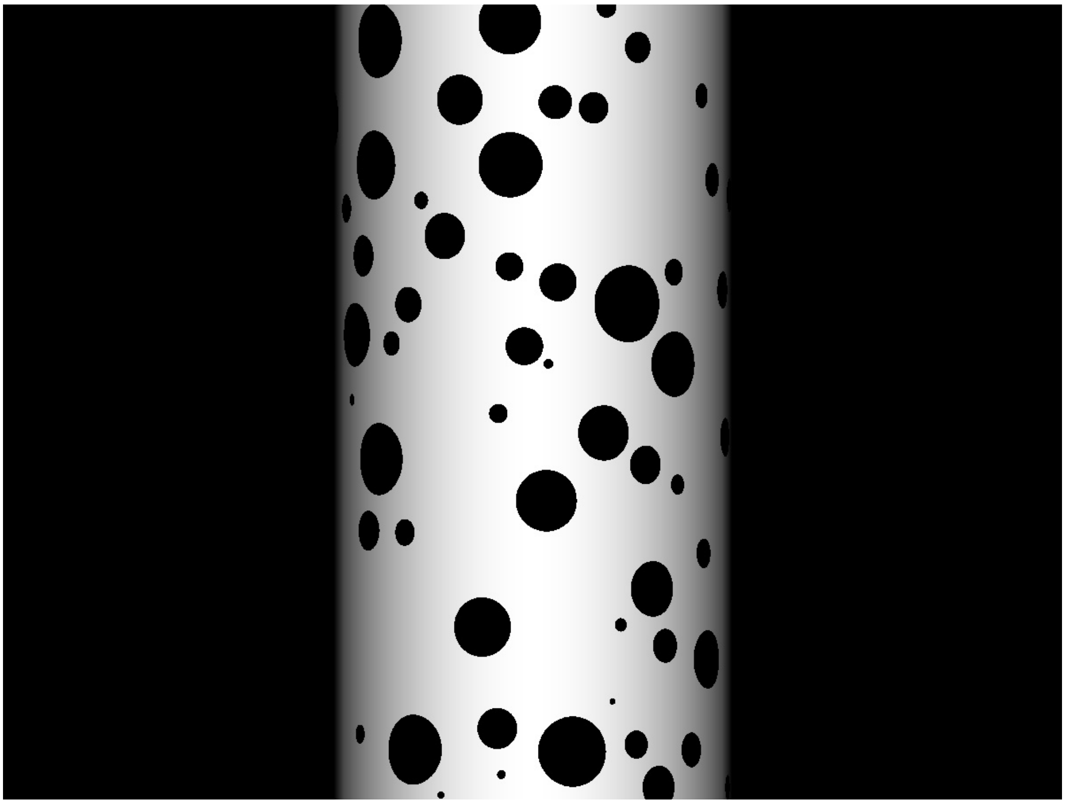


Figure 3.1. Example of the multiple-cue stimulus that was used in the Experiments 1 and 2. The stimulus shown here has consistent depth cues with a roughly circular cross section.

[3] Initially, we intended that the observers view the stimuli monocularly. In a preliminary test run, however, it turned out that it was very unpleasant to receive visual information on only one eye for session durations of 45 minutes or more. Thus, we decided to allow for binocular viewing in the final experiment.

3.2.2 Experimental design

Each observer carried out both a within-cue and a multiple-cue discrimination task in separate blocks. The within-cue discrimination task was performed for each cue (shading, texture, and motion) separately in order to obtain the single-cue reliabilities and the predicted weights for each respective depth cue. The standard stimulus cylinder always had a circular cross section in this task. In the multiple-cue task, the hemi-cylinder's depth was co-defined by all three visual cues. In the standard stimulus, one of the three cues was dissenting from the other two cues by 40, 80 or 120 pixels, respectively. The comparison stimulus comprised consistent depth cues.

3.2.3 Procedure

Within-cue discrimination task: For each observer, reliabilities of single cues were obtained from a within-cue discrimination task using a two interval, two alternative forced choice (2AFC) method. Each trial consisted of a sequential presentation of two stimuli that differed in their depths regarding the shading, the texture, or the motion cue, respectively. The other two dimensions were set to 0, that is they were non-informative with respect to the cylinder's depth. Each stimulus presentation lasted two seconds and consisted of one clockwise and one anticlockwise rotation for one second each. Between both presentation intervals, the screen was blank for one second. In the standard interval, the cylinder had a circular cross section, in the comparison interval, the cylinder's depth was larger than its width. The standard and the comparison stimulus were randomly assigned to the first or the second interval and the observer had to indicate the interval containing the deeper stimulus. The depth of the comparison stimulus was varied according to a 2-down-1-up adaptive staircase procedure targeting the 70.7% point of the psychometric function (Leek, 2001). For the first four reversals, step size was set to 25 pixels, from the 5th to the 10th reversal, step size was reduced to 10 pixels. The 70.7% point of the psychometric function was determined as the mean value of the last 6 reversals. The difference between the 70.7% point and the depth of the standard stimulus was regarded as a measure of single-cue discriminability (just noticeable difference, JND). Observers accomplished four sessions of the within-cue discrimination task. The first session that was not included in the analyses comprised an extensive training where feedback was given. During the other three sessions, each single-cue discriminability (shading, texture, and motion) was measured twice with randomly interleaved tracks. No feedback was given. The four testing sessions comprised 261 trials on average and lasted about 45 minutes each.

Multiple-cue task: To examine the impact of the different cues on the perceived depth of a multidimensional stimulus, the second part of the experiment consisted of a two interval 2AFC task using stimuli that contained all three depth cues that were only partially consistent with each other. In the standard stimulus, one depth cue was set to 280 pixels, the other two cues were set to 320 pixels. The comparison stimulus comprised consistent depth cues that were either clearly smaller or larger than the perceived depth of the standard stimulus. Again, the interval containing each stimulus was randomly determined and the observer had to indicate the interval containing the deeper stimulus. The depth of the comparison stimulus was varied either according to a 2-down-1-up or according to a 2-up-1-down adaptive staircase procedure, targeting the 70.7% or the 29.3% point of the psychometric function (Leek, 2001). Step size and threshold computation were equal to the within-cue discrimination task. In the first session, amounting to an extensive training, only stimuli with consistent depth cues were used and feedback was provided after each response. This session was not included in the analyses. In the subsequent three testing sessions without feedback, we measured the two above mentioned points of the psychometric functions for each deviating single cue, that is the depth cue that differed from the other two by 40 pixels. The resulting six tracks per session were randomly interleaved. On average, 269 trials were accomplished in each session which lasted about 45 minutes. The point of subjective equality (PSE) was determined for each condition (deviating shading, texture or motion cue, respectively) by averaging the three 70.7% and 29.3% points of the psychometric functions. The PSE was corrected for potential asymmetries by a linear interpolation using the thresholds that were determined in the training session on the basis of consistent depth cues. The JND was calculated using a similar procedure as for the within-cue discrimination task. Half of the observers started with the within-cue discrimination task, the others completed the multiple-cue task first. All eight sessions were accomplished within a period of three weeks.

To allow for a precise analysis of the influence of the dissenting cue on the perceived depth of the cylinder as a function of the difference between the dissenting and the two consistent cues, two additional multiple-cue conditions using a different amount of cue conflict were added subsequently to the experiment. Therefore, the observers accomplished six additional sessions during which one cue was set to 240 or 200 pixels, respectively, and the other two consistent cues indicated a depth of 320 pixels. All measurement details were equal to the multiple-cue task described above. For each condition, the two points of the psychometric functions were measured three times each with fully interleaved tracks, which were randomly allocated to the six testing sessions. The PSE for these new conditions was computed as described above. Each session comprised 258 trials on average and lasted about 45 minutes. The additional testing sessions were accomplished within a period of three weeks that started less than two weeks after the first eight sessions.

3.2.4 Parameter extraction and statistical analyses

The PSEs and JNDs from the within-cue discrimination task and the multiple-cue condition were used for the statistical analyses as well as for the modeling. All statistical analyses were carried out using R (version 2.4.1) or SPSS (version 14.0.2). An a priori alpha level of .05 was used for all statistical tests. The Huynh-Feldt procedure (Huynh & Feldt, 1976) was applied to correct for potential violations of the sphericity assumption in repeated-measures analyses of variance (ANOVAs) involving more than one degree of freedom in the enumerator. For each statistically significant effect in the ANOVAs, Cohen's *f* is reported as an effect size estimate (Cohen, 1988, p. 273 ff.).

3.2.5 Participants

Five naive observers (2 women, 3 men) who gave written informed consent and the author of this thesis participated in the experiment. Their ages ranged from 22 to 27 years (*M* = 24 years). All had normal or corrected-to-normal vision. The observers CM, JD and MG started with the within-cue discrimination task, AR, JM and RH completed the multiple-cue task first.

After the experiment, the naive observers were asked whether they had noticed that the cues in the multiple-cue condition indicated conflicting depths of the cylinder. None of them reported having noticed these discrepancies.

3.3 Results

3.3.1 Within-cue discrimination task

In a first step, the JNDs from the within-cue discrimination task were compared using an analysis of variance (ANOVA) with depth cue as within-subject factor. This analysis produced a significant main effect of cue, $F(2, 10) = 16.63$, $\varepsilon = .64$, $p < .001$, $f = 0.67$. Post hoc comparisons using Tukey's HSD procedure yielded significant differences between the JND of the shading cue (*M* = 42.92, *SD* = 24.39) and the texture (*M* = 97.69, *SD* = 45.36) and motion cue (*M* = 86.25, *SD* = 42.28), respectively (see Figure 3.3). The JNDs of the latter two cues did not differ significantly. Observers achieved the highest precision differentiating the depths of the cylinders when these were defined by the shading gradient. Thus, according to the MLE framework, the shading gradient that has the largest reliability should influence the perceived depth of the multiple-cue stimulus more strongly than both other cues.

3.3.2 Multiple-cue conditions

The empirical PSEs that were obtained from the multiple-cue task are depicted in Figure 3.2 as a function of the dissenting cue and its difference from the two consistent ones. A large influence of the dissenting cue on the perceived depth of the cylinder would be characterized by a displacement of the PSE into the direction of the bottom of the gray bars. A displacement into the direction of the bar's top would, on the other hand, indicate a high impact of the two consistent cues on the perceived depth of the cylinder.

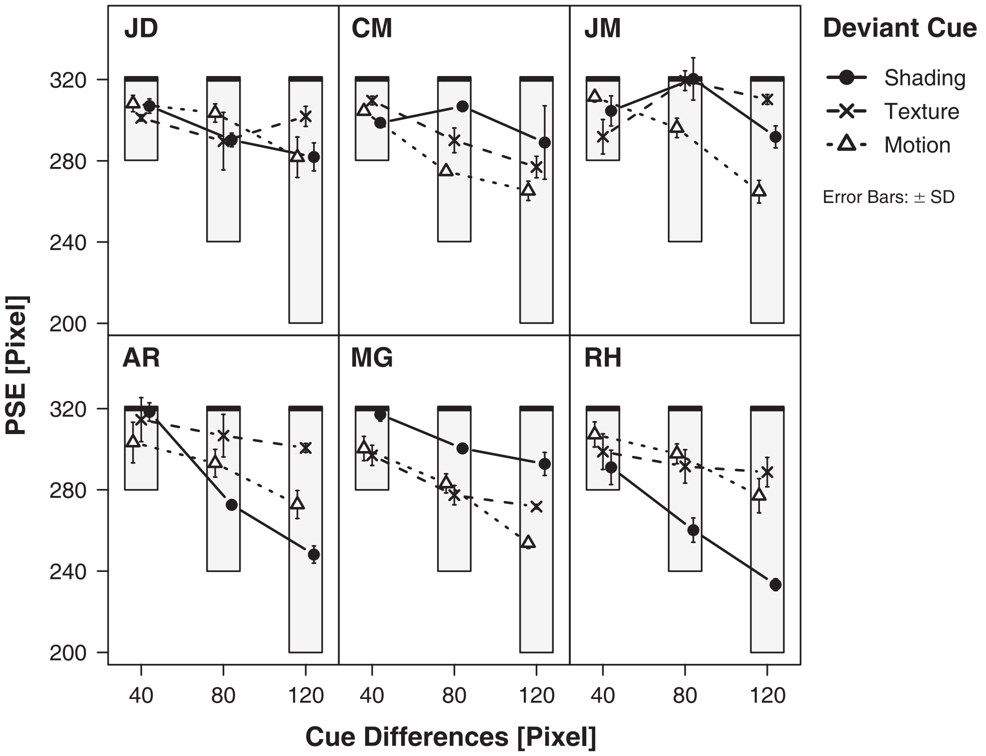


Figure 3.2. Points of subjective equality (PSE) as a function of the dissenting cue and its difference from the two consistent cues for each observer in Experiment 1. The gray bars depict the range between the dissenting cue which represents the bar's bottom and the two consistent cues that were always fixed to 320 pixels and correspond to the bar's top.

Only for observer RH, a large influence of the highly reliable shading cue could be observed. A similar tendency was obtained for observer AR, but exclusively for those conditions where the shading differed from the other two cues by at least 80 pixels. For all other observers, the perceived depth of the cylinder was largely determined by the two consistent cues instead of the dissenting shading cue. Observer JM for example seemed to completely override the shading cue when it differed from the other two cues by 80 pixels.

A 3 × 3 ANOVA on the empirical PSE using the within-subject factors dissenting cue and discrepancy between dissenting and consistent cues revealed a significant main effect of the latter factor, $F(2, 10) = 26.73$, $\varepsilon = .91$, $p < .001$, $f = 0.74$. Thus, the empirical PSE was shifted into the direction of the dissenting cue (q.v. Figure 3.5). However, the marginally significant interaction of both factors, $F(4, 20) = 2.63$, $\varepsilon = .79$, $p < .10$, $f = 0.27$, indicates that the displacement of the PSE tended to vary as a function of the dissenting cue type.

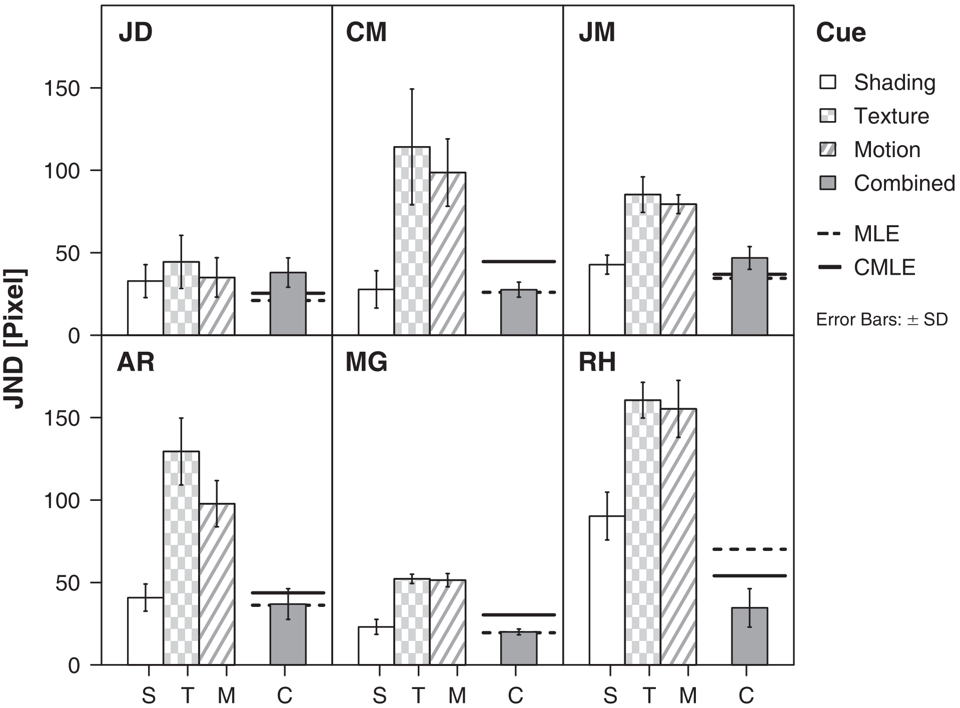


Figure 3.3. Just noticeable differences (JND) from the within-cue discrimination tasks and pooled values from the multiple-cue conditions for each observer in Experiment 1. The dashed line corresponds to the predicted value according to the MLE framework and the solid line denotes the prediction of the Bayesian model taking consistencies between cues into account (CMLE).

To examine whether the JNDs differed as a function of the dissenting cue or the amount of difference, a 3 × 3 ANOVA was conducted using the same factors as in the previous analysis. No significant main or interaction effects could be found. Thus, the JNDs remained relatively stable across the multiple-cue conditions. Figure 3.3 depicts the JNDs of each single cue and the pooled values from the multiple-cue conditions. Overall, the observers were at least as precise in estimating the depth of the multiple-cue stimulus as in the best single-cue condition. However, a marked reduction of the JND in the multiple-cue condition as predicted by ideal observer models (see section 2.4) was only examined for RH.

3.3.3 Model comparison

To examine whether cue interactions should be taken into account to explain the observer's performance in the multiple-cue condition, the free parameters of the Bayesian model (see section 2.5) were fitted to the empirical data. This was accomplished by minimizing the squared distance between predicted and empirical PSEs.

Figure 3.4 depicts the empirical PSEs that were obtained in the multiple-cue condition against the predictions of the two models. Table 3.1 shows the squared correlations (R^2) between the model's predictions and the empirical data for each observer. The simple MLE model provided an excellent fit only for RH's data and showed a satisfactory performance for the observers JD and AR. It was entirely unable to explain the data of CM, JM and MG. These observers did obviously not rely on single-cue reliabilities when judging the perceived depth of the cylinders in the multiple-cue condition. On the other hand, the Bayesian model including a coupling prior (CMLE model) was highly capable of explaining the data of CM and MG and showed a satisfactory fit to JM's data. These observers obviously took into account consistencies between cues when judging the cylinder's spatial depth. Additionally, the increase in the R^2 values for AR, suggests that she pursued a comparable strategy, albeit less pronouncedly.

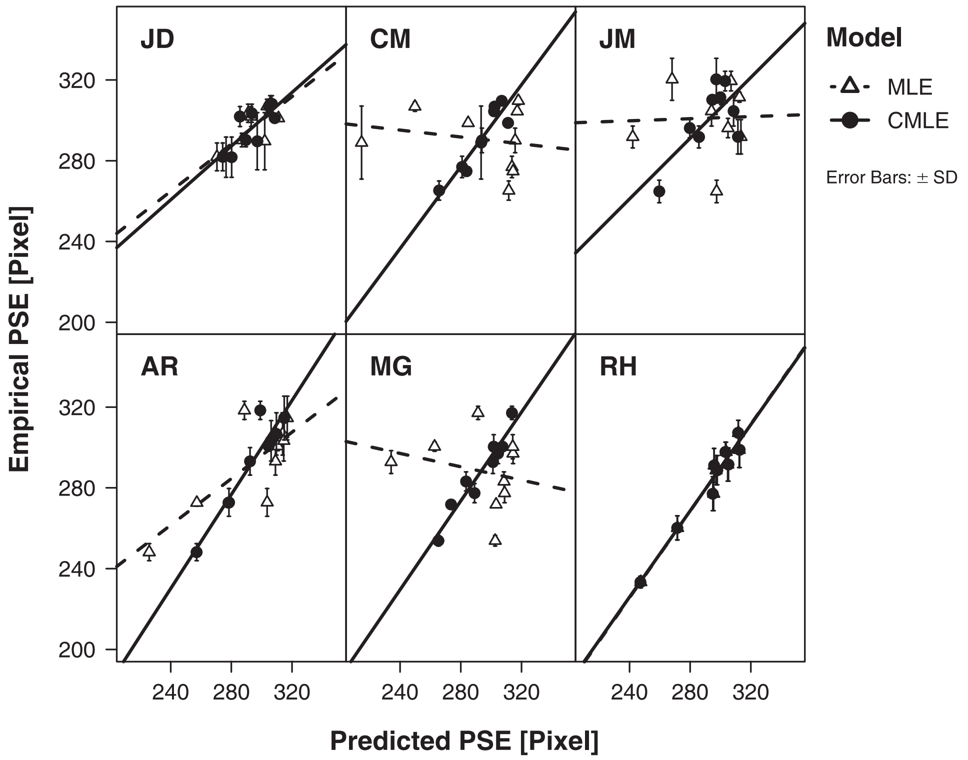


Figure 3.4. Scatterplot of predicted and empirical points of subjective equality (PSE) of Experiment 1 for both cue combination models. Solid and dashed lines represent regression lines for both models.

The main disadvantage of CMLE model concerns its complexity. For the current study, it has three free parameters whereas the MLE model has none. Thus, a comparison of both models only with respect to their fit to the empirical data might be inappropriate, as increasing model complexity always allows for a better fit to the empirical data. A fair comparison, however, can be achieved by relying on an information criterion that takes into account both model accuracy as well as model complexity. Raftery (1995) has proposed such a criterion based on Bayesian hypothesis testing, namely the Bayesian Information Criterion (BIC; q.v., Schwarz, 1978). Two models can be compared by determining their BIC values. The model with the smaller BIC value is preferable. We computed the BIC values for all observers and both competing models and finally subtracted the BIC values of the CMLE model from the corresponding values for the MLE model. Thus, a large positive BIC difference score would indicate superiority of the CMLE model. A large negative score, on the other hand would indicate superiority of the MLE model. As can be seen from Table 3.1, even when taking model complexity into account, the CMLE model was superior for observers CM, AR and MG. Both models were equivalent for observer JM and only for the observers JD and RH, the MLE model did show a larger degree of plausibility.

Table 3.1. Fit of the MLE and the CMLE model as indicated by R^2 values for the PSE data of Experiment 1

Observer	R^2_{MLE}	R^2_{CMLE}	ΔBIC
JD	.59	.60	-6.49
CM	.04	.87	11.46
JM	.00	.49	-0.54
AR	.61	.89	4.93
MG	.06	.93	16.64
RH	.96	.96	-6.65

Note. A model comparison was achieved by computing the difference of the Bayesian Information Criterion (*BIC*) scores of both models.

How did the coupling prior of the Bayesian model contribute to the somewhat better fit of this model to the empirical data? This question can be answered by inspecting the pooled data across observers along with the predictions of both models (see Figure 3.5). The observers did obviously reduce the weight of the shading cue beyond the predictions of the MLE model while increasing the texture and motion weighting. Across observers, the CMLE model captured this differential weighting more adequately than the MLE framework.

From the predicted PSEs of both models, the average change in single-cue weights can be computed for each observer (see Table 3.2). An ANOVA on the weight changes using depth cue as within-subject factor revealed a significant main effect, $F(2, 10) = 7.72$, $\varepsilon = .59$, $p < .05$, $f = 1.24$. Post hoc comparisons using Tukey's HSD procedure yielded only a marginally significant difference between weight changes of the shading and the motion cue.

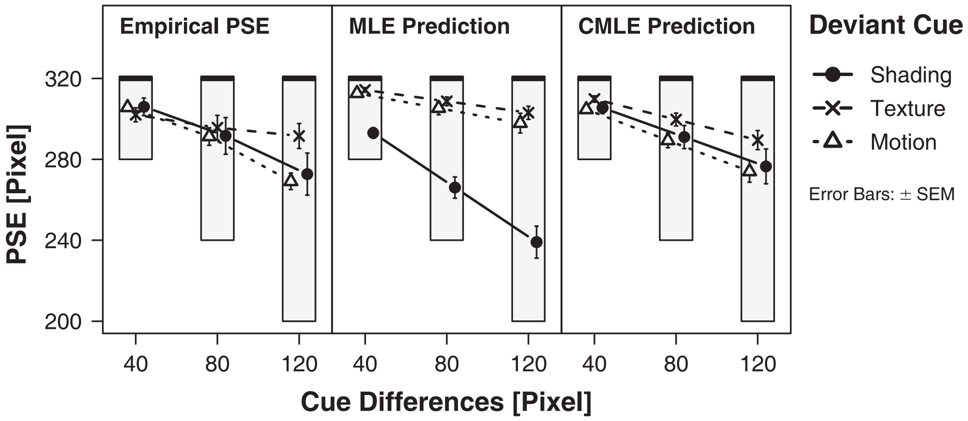


Figure 3.5. Average points of subjective equality (PSE) as a function of the dissenting cue and its difference from the two consistent cues in Experiment 1. In the left panel, the empirical PSEs are depicted; the middle and right panel show the predictions of the MLE and the CMLE model, respectively. The gray bars depict the range between the dissenting cue which represents the bar's bottom and the two consistent cues that were always fixed to 320 pixels and correspond to the bar's top.

Obviously, the shading cue did not receive its appropriate weight when the texture and motion information consistently indicated a different depth of the stimulus. This was clearly the case for observers CM, MG, and JM, and also evident for AR. On the other hand, texture and motion cues to the cylinder's depth

Table 3.2. Average changes in single-cue weights that were induced by the coupling prior in Experiment 1

Observer	Δw_S	Δw_T	Δw_M
JD	-.03	.06	-.03
CM	-.66	.27	.38
JM	-.36	.05	.31
AR	-.26	.05	.21
MG	-.56	.25	.31
RH	.00	-.01	.00
Across observers	-.31 (.27)	.11 (.12)	.20 (.17)

Note. The subscripts S, T and M denote the shading, the texture, and the motion cue, respectively. Average values are depicted in the bottom row with the standard deviation in brackets.

received a larger weight than expected on the basis of their single-cue reliability, when they were dissenting from the other two consistent cues. Only observers JD and RH seemed to ignore the pairwise consistencies between the depth cues in the multiple-cue condition, as their cue weighting did not differ appreciably from the within-cue condition. For these two observers, the MLE model explained their empirical data fairly well.

Another way to evaluate the adequacy of both models is an analysis of the predicted JNDs in the multiple-cue condition. Figure 3.3 depicts the predictions of both models as dashed and solid lines behind the bar representing the pooled JNDs of the multiple-cue conditions. Overall, no clear advantage of either model could be observed. But interestingly, only the CMLE model was able to predict a noticeably reduced threshold of observer RH in the multiple-cue situation although his PSEs were well explained by an MLE integration scheme.

3.4 Discussion

The present study examined whether temporary consistencies between visual depth cues influence a combined depth estimate above and beyond the influence of single-cue reliabilities. For this purpose, visual displays of hemi-cylinders with an elliptical cross section were constructed whose depth was defined by shading, texture, and motion information. All cues could be varied independently from each other, thus allowing for arbitrary consistencies between cues. To examine whether these cue consistencies affect their integration, predictions of two cue integration models were compared to the empirical data. One model solely relied on single-cue reliabilities when predicting the observer’s responses (MLE model). The other model used differential coupling priors in a Bayesian framework to allow for a flexible reweighting as a function of cue consistencies (CMLE model).

3.4.1 The power of shading

In the within-cue discrimination task, all observers achieved the highest precision in estimating the cylinder's depth on the basis of the shading gradient. This pattern of differential discrimination thresholds was rather unexpected because smooth shading is generally a weak cue to an object's shape (e.g., Bülthoff & Mallot, 1988; Erens, Kappers, & Koenderink, 1993; Todd, Norman, Koenderink, & Kappers, 1997). Some empirical findings, however, may help to clarify this discrepancy. First, we used an object with a comparably simple convex shape. Liu and Todd (2004) demonstrated that observers have a strong perceptual bias to interpret images as convex rather than concave. Additionally, they showed that estimated and simulated depths of ellipsoidal stimuli were substantially correlated only for convex surfaces (q.v., Todd & Mingolla, 1983). Moreover, Mingolla and Todd (1986) reported a high correspondence between shape-from-shading judgments and an actual ellipsoid's shape at least when the object was of comparably low eccentricity, as were the cylinders in the current study. The shading gradient of all displays was produced by a laminar light source directed parallel to the observer's line of sight in both experiments of the present study. Thus, typical changes in lighting direction that often influence shape-from-shading judgments (Bülthoff & Mallot, 1988; Curran & Johnston, 1996; Mingolla & Todd, 1986) did not occur in our study. Taken together, the performance of a human depth-from-shading algorithm seems to be optimized for conditions that closely match the stimuli of the current study.

Another issue might also help to explain the high precision of the depth-from-shading estimates. We chose to normalize the shading gradients in both experiments to force the observers to concentrate on depth perception of the object instead of focusing on the observer-object-distance (see Koenderink, van Doorn, Christou, & Lappin, 1996). In principle, this normalization impeded an estimation of the distance between the light source and the cylinder's surface. On the other hand, normalized shading gradients might have encouraged our participants to use a simple brightness heuristic to estimate the hemi-cylinders' depths on the basis of the shading cue. Such a heuristic would simply state that flat cylinders are brighter than deep ones. As human difference thresholds are very small for brightness changes under the photopic conditions that were used in the current study (Pöppel & Harvey, 1973) this may explain the high precision of depth-from-shading judgments.

3.4.2 Coupling between cues

Given that the stimuli of the current study obviously allowed for a very precise evaluation of the hemi-cylinder's depth based on the shading cue, it is all the more astonishing that this cue was heavily downweighted when judging the depth of displays with consistent texture and motion cues. Moreover, observers tended to increase the weight of the texture or motion cue, respectively, when these two differed from each other. An explanation for this result, which seems to be counterintuitive at the first glance, may be related to prior experiences with these cues as potential indicators of an object's depth. In naturalistic scenes, texture and motion cues are always strongly correlated, whereas the shading cue is vulnerable to changes in an object's illumination. An observer taking into account this prior knowledge would thus be prone to put a larger weight on consistent texture and motion cues while suppressing the potentially misleading depth-from-shading estimate. Conversely, whenever discrepancies between texture and motion cues occur, both are supposed to carry independent information about the visual scene. To fully use this non-redundant information it might be useful to heighten the weight assigned to a dissenting cue (Atkins et al., 2003).

In the multiple-cue condition, JNDs were at least as small as in the best single-cue condition. However, it seems unlikely that observers solely relied on the most reliable cue as a simple heuristic to estimate the cylinder's depth in the multiple-cue condition. Figure 3.5 clearly reveals that all cues contributed to estimated depth when being available to the observer.

3.4.3 Potential limitations

The results of the current study indicate that prior expectations regarding cue consistencies affect sensory integration during depth perception. However, some limitations should be acknowledged: First, the standard stimulus in the within-cue discrimination task had a constant depth of 300 pixels whereas four different depths (200, 240, 280, and 320 pixels) of each cue were used in the multiple-cue conditions. Empirical evidence suggests that the JNDs of the different cues might vary as a function of cylinder depth. For example, Knill (1998b) has shown that the reliability of texture cues varies as a function of the slant of planar surfaces. Comparable results were reported by Knill and Saunders (2003) as well as by Rosas and colleagues (2004). It is unclear whether these results which were obtained in a slant discrimination task also apply to depth estimates of cylinders with an elliptical cross section, but it might be speculated that the poor performance of the MLE model can be partially explained by the fact that weights varied as a function of cylinder depth. Two arguments can be put forward against this speculation (see Appendix B): First, when taking the depth estimation task as a special case of a slant discrimination experiment, the maximum difference between tangential angles to the cylinder's surface in the single- and the multiple-cue condition varied between -1.85° and 11.54° across all possible fixation points (Appendix B 1.1). Given these comparatively small differences, it seems unlikely that JNDs of single cues have differed substantially across the cylinder depths studied here. Indeed, an empirical examination using one observer of the original study (MG) revealed that JNDs of single cues were relatively stable across different cylinder depths of the standard stimulus in the within-cue discrimination task (Appendix B 1.2).

A second limitation concerns the measurement of the single-cue reliabilities. This was achieved by varying one depth cue while fixing the other two dimensions at 0. This procedure was utilized to maximize the similarity of single-cue and multiple-cue conditions. One could argue, however, that in this case the two fixed dimensions had carried depth information indicating flatness.

Moreover, these stimuli did not correspond to an object that could be constructed outside the virtual world and therefore they looked somewhat artificial. This concern will be followed up in Experiment 2 where different single-cue conditions were added to the experimental design to remedy this limitation.

3.4.4 Conclusions

Taken together, the superiority of the CMLE over the MLE model in predicting the empirical PSEs was demonstrated for four of six observers in the current study. Even when taking model complexity into account, a clear advantage of the CMLE model was found for half of the observers. Thus, depth estimates were clearly affected by trialwise interactions between visual depth cues. This finding extends corresponding results from cross-modal studies that relied on consistencies between visual and haptic cues (Atkins et al., 2001, 2003; Ernst et al., 2000; q.v., Jacobs, 2002b). Furthermore, it was demonstrated that incidental consistencies between the cues were sufficient to induce a flexible reweighting. This happened spontaneously, neither extensive training nor conscious detection of cue interactions was required.

4. Experiment 2

4.1 Introduction

Experiment 1 showed that in addition to single-cue reliabilities, prior expectations about consistencies between visual cues affected depth perception. The second experiment sought to replicate and extend these findings. To remedy potential limitations of Experiment 1 (see 3.4.3), several improvements of the experimental procedure were implemented. First, single-cue reliabilities were measured separately for each depth that was used in the multiple-cue task to allow for precise predictions of the MLE model across varying reliabilities. Second, in addition to the single-cue conditions that were used in Experiment 1, a new single-cue task was developed. Stimuli were constructed that were defined solely by one of the three depth dimensions without including the other two cues at all (see Jacobs, 1999, for a similar procedure). This allowed for an examination of the appropriateness of the within-cue discrimination task that was employed in Experiment 1. Third, all three cues were varied orthogonally in the multiple-cue condition across a wide range of depths. Thus, the adequacy of both sensory integration models could be tested for stimuli with fully consistent, fully discrepant, and partly discrepant depth cues. For economic reasons, a depth-matching instead of a two interval 2AFC task was used in this experiment.

4.2 Method

4.2.1 Stimuli and apparatus

The stimuli used in this experiment were the same hemi-cylinders with an elliptical cross section as in Experiment 1. In addition to the single- and multiple-cue conditions described above, we employed stimuli whose depth was solely determined by one of the three respective visual cues. The shading-only stimulus consisted solely of the shading gradient with no texture mapped onto the cylinder's surface. In the texture-only condition, white circular texture elements were mapped onto a uniformly black surface, that is, the cylinder's background seemed to be invisible. The same parameters as in Experiment 1 were used for the mapping algorithm. Neither shading nor motion information was present in this scene. The motion-only condition was realized by wrapping a two-dimensional sheet with 1000 × 600 pixels around an otherwise invisible cylinder. Between 180 and 220 white dots were randomly placed onto this sheet, their radius was randomly chosen to equal 0.5 or 1 pixel. The distance between two adjacent dots was at least 10 pixels. The motion information was constructed using the same algorithm and the same velocities as in Experiment 1[4]. Screenshots of all newly introduced conditions are depicted in Figure 4.1.

Stimuli were computed in real time on the same computer and displayed on the same 19" monitor as in Experiment 1 with a resolution of 800 × 600 pixels and a color depth of 32 bits. A second monitor (19" Scott 995) was placed directly beside the display where the stimuli were shown. This monitor was connected to a second PC and was used to display a top view of the outline of a hemi-cylinder, that is, a hemi-ellipse. The hemi-ellipse was plotted white on black background. The horizontal diameter of this ellipse was identical to the cylinder's width on the other monitor and was placed 50 pixels below the screen top. The vertical radius, on the other hand, could be increased or decreased in steps of 1 pixel by use of the cursor keys. Using this method, cylinder depths from 0 to 600 pixels could be produced. Both computers were placed in a windowless dimly lit room and stimuli were viewed from a distance of 1 m. To allow for a comparison to

[4] Strictly speaking, the motion-only condition did not solely contain motion cues to the cylinder's depth but additionally comprised texture compression information because of the light point's density gradient. Nevertheless, this stimulus might be appropriately used to determine the single-cue reliability of the motion cue, because: 1) In general, density gradients are a weak cue to an object's shape (e.g., Knill, 1998a) and it has been frequently reported that density cues only receive a very small weight compared to other textural information in visual scenes (e.g. Blake et al., 1993; Buckley, & Frisby, 1993; Buckley et al., 1996; Cumming, Johnston, & Parker, 1993; Cutting & Millard, 1984; Knill, 1998c). 2) We tried to minimize the density information by mapping only a comparatively small number of light points onto the cylinder's surface. 3) Other researchers have also made this assumption and did successfully use similar conditions to measure the reliability of the motion cue to an object's depth (e.g. Jacobs, 1999).

Experiment 1, stimuli were viewed binocularly. As in Experiment 1, the displayed image of the cylinder subtended 7.8° of visual angle in the horizontal and 15.6° in the vertical dimension. The vertical midlines of both display screens were separated by 24.3° of visual angle.

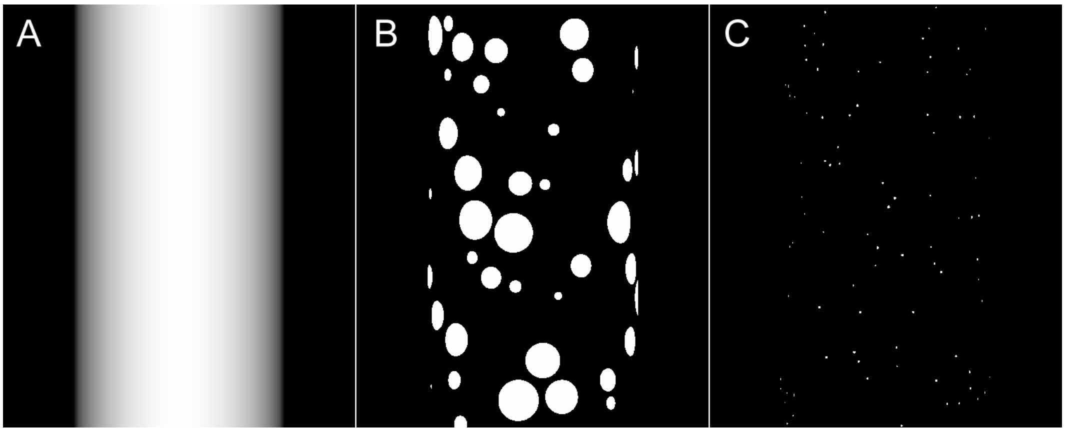


Figure 4.1. Examples of the stimuli that were additionally used in Experiment 2 to determine the single-cue reliabilities. The depth of the cylinder shown in panel A) could only be estimated on the basis of the shading gradient. Panel B) shows a cylinder whose depth is solely defined by the texture compression and panel C) depicts one frame of the motion condition. All stimuli shown here have a roughly circular cross section.

4.2.2 Experimental design

Single-cue PSEs and JNDs were determined using a magnitude estimation method. The stimuli used in this task either contained only one informative visual depth cue (shading, texture, or motion), which indicated a depth differing from flatness (cf., Experiment 1) or solely comprised one of the three visual depth cues (see Figure 4.1). In the multiple-cue task, the cylinder's depth was defined by a combination of shading, texture, and motion cues. The depth of these cues was varied orthogonally and each cue could take one of three depth values (200, 300, or 400 pixels). Thus, twenty-seven different combinations of fully consistent, partly consistent, or entirely discrepant depth cues were realized.

4.2.3 Procedure

In each trial, the hemi-cylinder with the elliptical cross section was displayed on one monitor and the observer had to adjust the outline of the hemi-ellipse that was displayed on the other monitor to match the hemi-cylinder's depth. The initial height of the hemi-ellipse was randomly drawn from a rectangular distribution ranging from 0 to 300 pixels. The observer finished a trial by pressing the "Enter" key. The next trial started approximately 2 seconds later. No time limit was specified for the adjustments. The observers needed 11.773 s on average (*SD* = 4.175 s) for each trial. In the single-cue task, the hemi-cylinder's depth was defined by one cue only. In the multiple-cue task, it was defined by a combination of all three cues.

Single-cue task: Two types of single-cue conditions were utilized to compute the PSEs and JNDs of each cue. The single-cue conditions were either the same as in Experiment 1, where only one cue was set to values different from 0 while the other two cues indicated flatness (one cue relevant), or they comprised stimuli whose depth was solely defined by one of the three visual depth cues (one cue alone, see Figure 4.1). For each condition and cue, three depths of the cylinder were realized that corresponded to the depths that were used in the multiple-cue task: 200, 300, and 400 pixels. Each combination was repeated ten times, resulting in a total of 300 trials[5].

Multiple-cue task: In the multiple-cue task, the cylinder's depth was simultaneously defined by shading, texture, and motion cues. Each depth cue was set to 200, 300, or 400 pixels independently of the other two depth cues. Thus, twenty-seven different combinations were realized. Three of these combinations comprised consistent depth cues, six combinations fully discrepant depth cues, and in the remaining combinations one cue differed from the other two, the latter being consistent. Each combination was repeated ten times resulting in a total of 270 trials.

[5] Two additional depths of 100 and 500 pixels were used in the single-cue condition to prevent the observers from anticipating depth values. These conditions were not analyzed because they did not correspond to the depths that were used in the multiple-cue task.

Before the single- and the multiple-cue tasks, a training block familiarized the observers with the task. The hemi-cylinder that was displayed in the training did either match one of the single-cue conditions or a multiple-cue condition with consistent depth cues. The depth of the cylinder was randomly varied from trial to trial according to a rectangular distribution that ranged from 100 to 500 pixels. After the observer had adjusted the hemi-ellipse on the other monitor and confirmed her adjustment by pressing the "Enter" key, feedback was provided. If the observer's adjustment differed from the actual depth of the cylinder by less than 13 pixels, the trial was designated to be correct; otherwise it was defined to be wrong. Feedback was given by printing the words "correct" or "wrong" at the bottom of the monitor that displayed the hemi-ellipse. Additionally, a hemi-ellipse that corresponded to the actual depth of the hemi-cylinder was overlaid onto the observer's adjustment. The colour of the overlay as well as the colour of the written feedback was green for correct and red for wrong trials. Feedback was displayed for two seconds, afterwards the new trial started. This feedback was only given during training and was not shown for test trials.

Observers participated in the experiment for 4 testing sessions that usually occurred on different days. Each session lasted about one hour. The first session comprised 210 training trials (30 of each condition). All trials of the single- and the multiple-cue conditions were randomly assigned to the remaining three testing sessions (190 trials per session). Additionally, each of these testing sessions started with 70 training trials (10 of each condition) to allow for a short practice of the task and a potential refreshment of task strategies. All 4 sessions were accomplished within two weeks.

4.2.4 Parameter extraction and statistical analyses

For each single- and multiple-cue condition, two parameters were calculated from the empirical data. First, the PSE was computed as the mean of all ten adjustments that were accomplished within each condition. Additionally, the JND of the respective condition was calculated as the standard deviation of these adjustments.

All statistical analyses were carried out using R (version 2.4.1) or SPSS (version 14.0.2). An a priori alpha level of .05 was used for all statistical tests. As in Experiment 1, the Huynh-Feldt procedure (Huynh & Feldt, 1976) was applied to correct for potential violations of the sphericity assumption in repeated-measures ANOVAs involving more than one degree of freedom in the enumerator. Cohen's f is reported as an effect size estimate for each statistically significant effect in the ANOVAs (Cohen, 1988, p. 273 ff.).

4.2.5 Participants

The sample consisted of eight observers (5 women, 3 men) who had not participated in Experiment 1. All gave written informed consent after being told that participation was voluntary and that they could withdraw from the experiment at any time. Their age ranged from 21 to 32 years (M = 24 years). All had normal or corrected-to-normal vision and all were naive to the purposes of the experiment. After the experiment, observers were asked whether they had noticed discrepancies between the cues in the multiple-cue condition. None of them reported having noticed these conflicts.

4.3 Results

4.3.1 Single-cue conditions

In a first step, it was tested whether both methods that were used in the single-cue conditions produced comparable results. To this end, separate 2 × 3 × 3 ANOVAs were conducted on the PSE and the JND using the type of the single-cue task (one cue relevant or one cue alone), the cue (shading, texture, or motion) and the depth of the stimulus (200, 300, or 400 pixels) as within-subject factors. With respect to the PSEs, significant main effects of cue, $F(2, 14) = 5.79$, $\varepsilon = .62$, $p < .05$, $f = 0.24$, and depth, $F(2, 14) = 47.08$, $\varepsilon = .65$, $p < .001$, $f = 1.05$, were obtained along with a significant interaction of both factors, $F(4, 28) = 10.71$, $\varepsilon = .64$, $p < .001$, $f = 0.21$. As can be seen from Figure 4.2, the PSEs increased linearly with the depth of the stimulus. Moreover, the shading cue led to a more pronounced increase as compared to both other cues. Most importantly, the estimation method did not produce any significant main or interaction effect, thus, both methods produced comparable results.

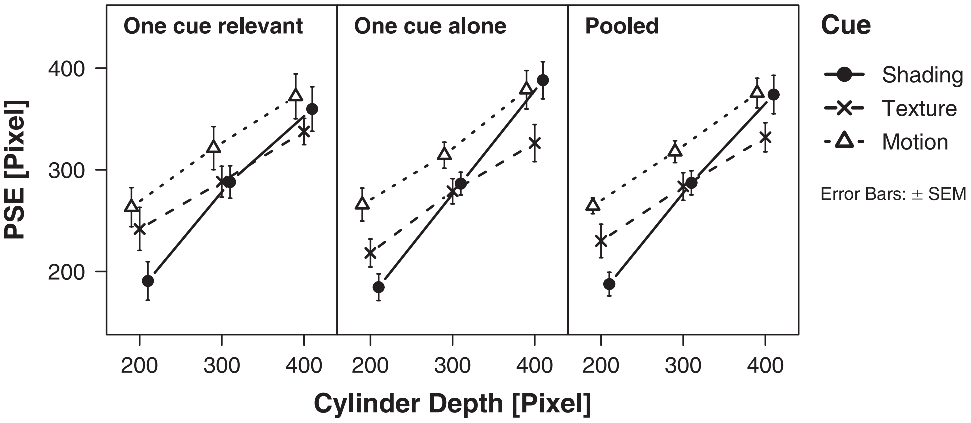


Figure 4.2. Points of subjective equality (PSE) as a function of cue and cylinder depth for both estimation methods (one cue relevant vs. one cue alone) in the single-cue conditions of Experiment 2. Additionally, pooled values across both methods are depicted.

With respect to the JNDs, the ANOVA yielded significant main effects of cue, $F(2, 14) = 4.56$, $\varepsilon = .98$, $p < .05$, $f = 0.27$, and depth, $F(2, 14) = 4.30$, $\varepsilon = 1.00$, $p < .05$, $f = 0.18$. Overall, the texture cue tended to be less reliable than the other two cues and JNDs were largest when the hemicylinder had a circular cross-section (Figure 4.3). Additionally, a marginally significant main effect for the method factor was obtained, $F(1, 7) = 3.95$, $p < .10$, $f = 0.15$, which suggests that the single-cue condition using one cue alone was slightly easier (i.e. it produced lower JNDs) than the other method using one relevant cue. No significant interaction was obtained for the estimation method. Taken together, both methods were largely comparable. Thus, data from both tasks was pooled for each cue × cylinder depth combination (see Figure 4.4 for a detailed illustration of the JNDs of each observer). These values were used in the modeling section below (section 4.3.3).

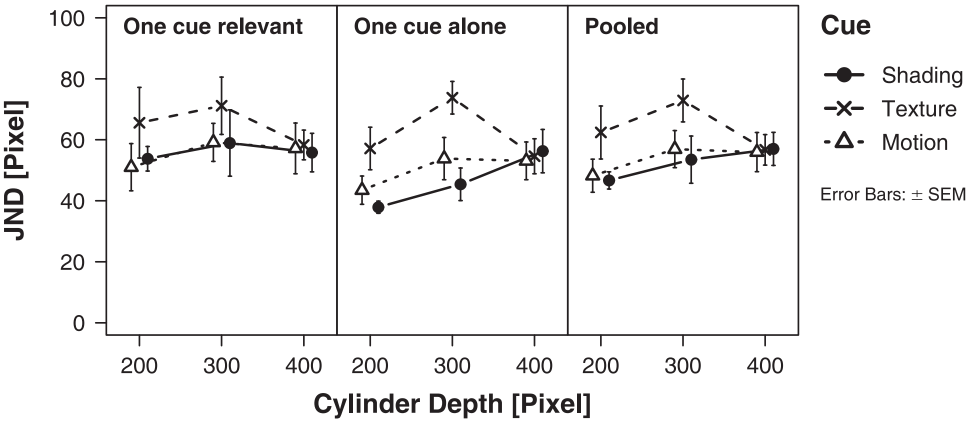


Figure 4.3. Just noticeable differences (JND) as a function of cue and cylinder depth for both estimation methods (one cue relevant vs. one cue alone) in the single-cue conditions of Experiment 2. Additionally, pooled values across both methods are depicted.

4.3.2 Multiple-cue conditions

To examine whether different depths of the shading, texture, and motion cue respectively influenced the perceived depth of the multiple-cue stimulus across observers, a 3 × 3 × 3 ANOVA on the empirical PSEs was conducted. The within-subject factors represented the three different cues with the factor levels 200, 300, and 400 pixels. Significant main effects for the shading, $F(2, 14) = 10.22$, $\varepsilon = .51$, $p < .05$, $f = 0.60$, texture, $F(2, 14) = 11.73$, $\varepsilon = .71$, $p < .01$, $f = 0.19$, and motion cue were obtained, $F(2, 14) = 7.24$, $\varepsilon = .64$, $p < .05$, $f = 0.31$. Thus, each of these dimensions influenced the perceived depth of the stimulus. Moreover, a marginally significant interaction of texture and motion depths, $F(4, 28) = 2.63$, $\varepsilon = 1.00$, $p < .10$, $f = 0.07$, might indicate a coupling of both cues (see 4.3.3 for more detailed analyses regarding this issue).

A comparable 3 × 3 × 3 ANOVA was also carried out on the JNDs of the multiple-cue conditions. No significant main or interaction effect could be found. Thus, the JNDs remained relatively stable across the multiple-cue conditions and they were pooled for each observer to allow for a comparison to the single-cue JNDs (Figure 4.4). Overall, multiple-cue JNDs resembled the JNDs from the single-cue conditions. A marked reduction as predicted by ideal observer models (see section 2.4) was not found for any of the observers.

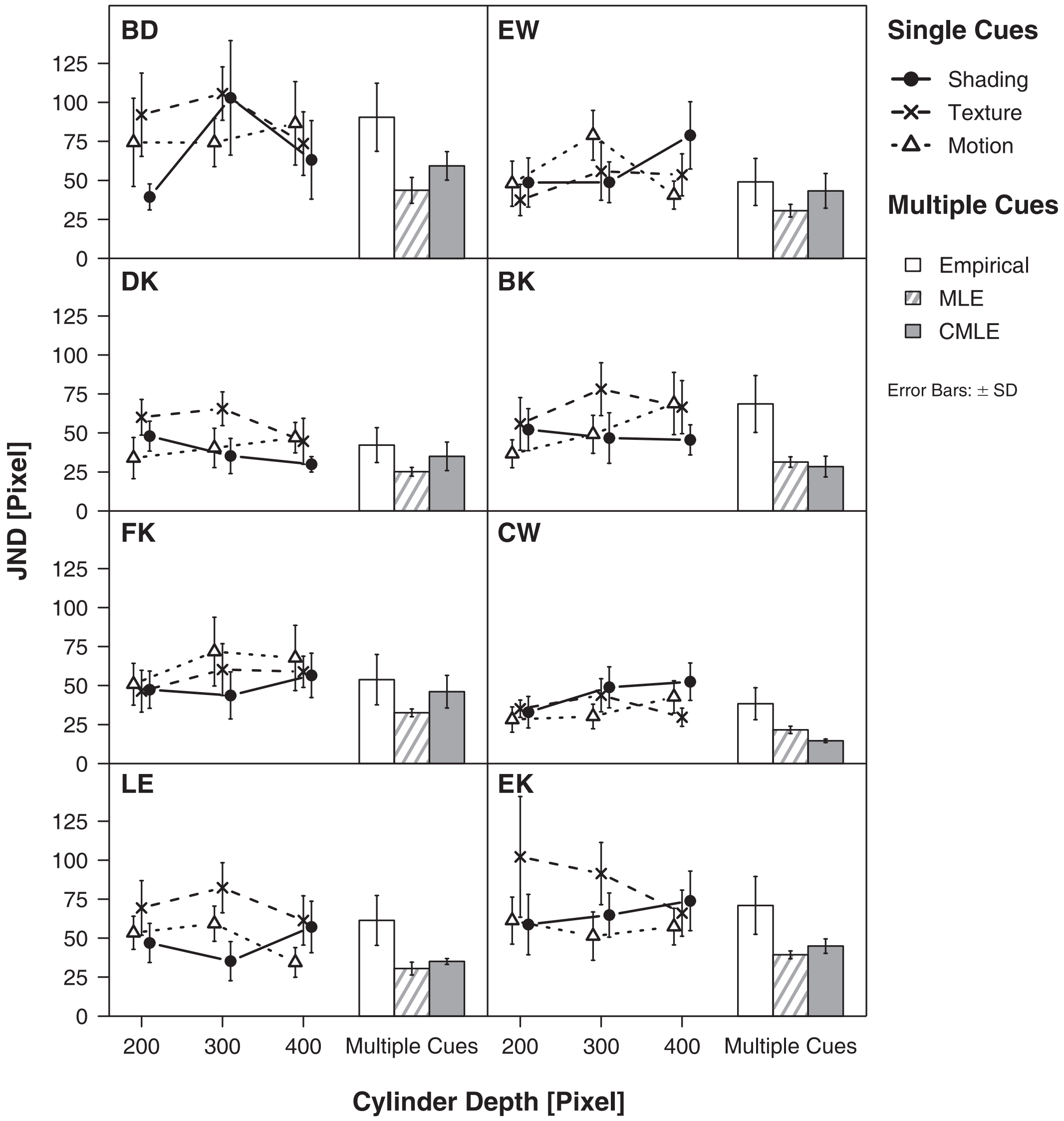


Figure 4.4. Just noticeable differences (JND) as a function of cue and cylinder depth for each observer in Experiment 2. Additionally, pooled values from the multiple-cue condition as well as the predictions of the MLE and the CMLE model are depicted.

4.3.3 Model comparison

In order to compare the MLE model to the Bayesian approach that takes into account prior expectations regarding the coupling of cues, predicted PSEs for the multiple-cue condition were calculated for each model. The PSEs and JNDs from the single-cue conditions of each observer were utilized for this purpose to take into account variations of these values across different depths of the cylinder (see Jacobs, 1999, for a similar procedure). The free parameters of the Bayesian model (see section 2.5) were fitted to the empirical data by minimizing the squared distance between predicted and empirical PSEs.

Table 4.1. Fit of the MLE and the CMLE model as indicated by R^2 values for the PSE data of Experiment 2

Observer	R^2_{MLE}	R^2_{CMLE}	ΔBIC
BD	.83	.89	2.45
EW	.41	.86	28.10
DK	.89	.95	10.29
BK	.78	.81	-5.69
FK	.71	.86	10.34
CW	.52	.72	4.55
LE	.23	.44	-1.28
EK	.47	.66	2.12

Note. A model comparison was achieved by computing the difference of the Bayesian Information Criterion (*BIC*) scores of both models.

The fit of both models to the empirical data can be evaluated using Table 4.1 and Figure 4.5. The MLE model provided a satisfactory fit for the observers BD, DK, BK, and FK but was outperformed by the CMLE model for all observers except BK and LE, as indicated by the differences in the Bayesian Information Criterion that additionally accounts for model complexity (Raftery, 1995; see section 3.3.3). For observer LE, however, the fit of both models was poor. Maybe she unpredictably switched her strategy between the single-cue and the multiple-cue conditions. The largest gain in the prediction of the empirical PSEs by the CMLE model was achieved for observer EW. As can be seen from Figure 4.5, this benefit holds for all types of multiple-cue stimuli. This is, to a lesser degree, also true for observer FK.

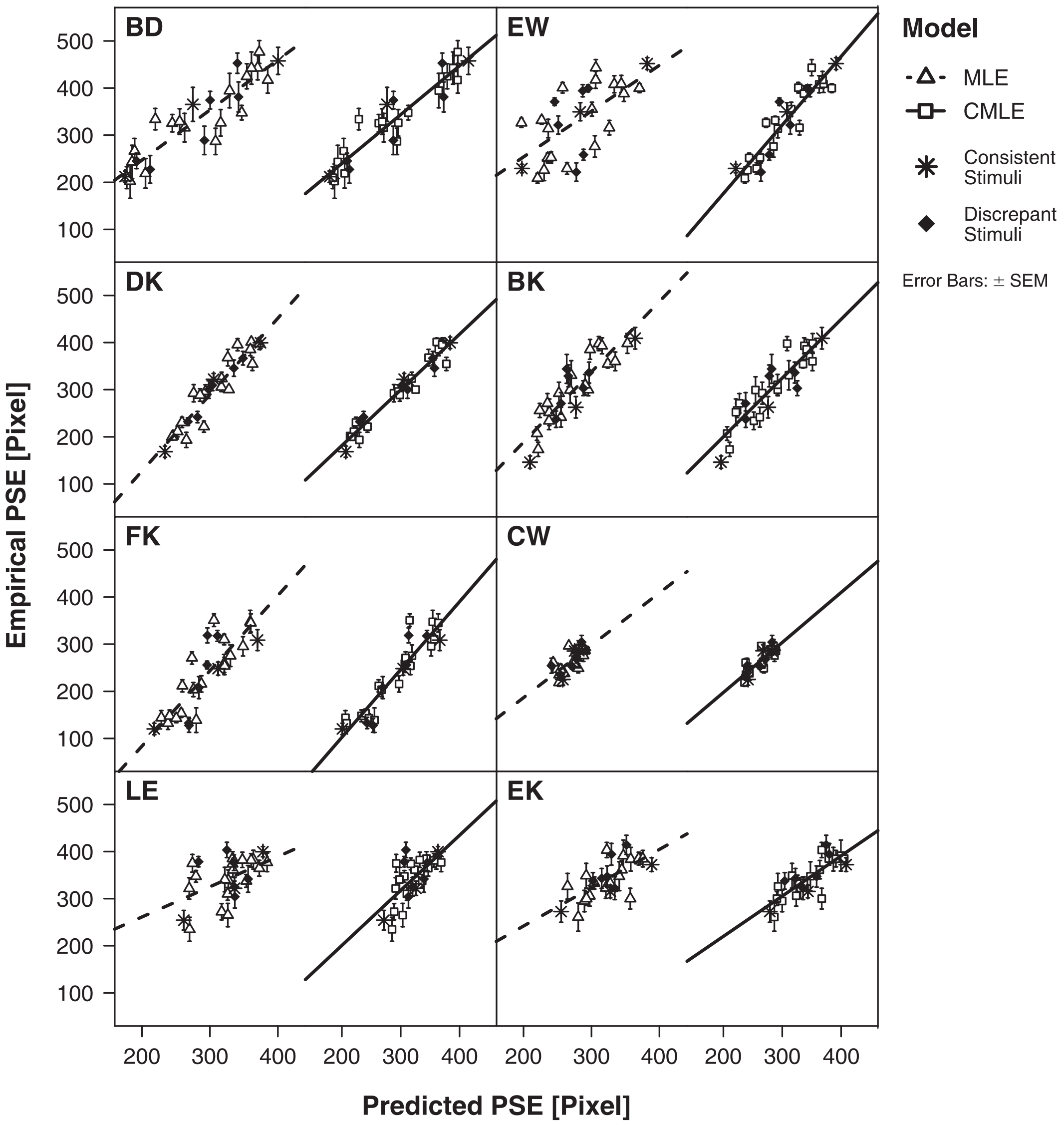


Figure 4.5. Scatterplot of predicted and empirical points of subjective equality (PSE) of Experiment 2 for both cue combination models. Stimuli with fully consistent and entirely discrepant depth cues are displayed separately. Solid and dashed lines represent regression lines for both models.

Although the coupling prior of the Bayesian model was optimized to predict the empirical PSEs, it would be interesting to examine whether the JNDs of the multiple-cue condition also follow the model's predictions. Empirical and predicted JNDs are depicted in Figure 4.4 for each observer. Although both models did not fully capture the variations of the JNDs across observers, it seems that the CMLE model provided a better fit than the MLE model, which consistently underestimated the JNDs. The predicted JNDs of the Bayesian model were in close agreement with the empirical JNDs especially for the observers EW, DK, and FK.

4.4 Discussion

Using a comparable design as Experiment 1, the present study examined whether in addition to single-cue reliabilities, temporary consistencies between shading, texture, and motion cues to visual depth influence a combined depth estimate. Several improvements were realized to remedy potential limitations of Experiment 1 (see 3.4.3). These were primarily related to the measurement of single-cue reliabilities as well as to the combination of single cues in the multiple-cue conditions.

4.4.1 Comparability of single-cue conditions

In Experiment 1, single-cue reliabilities were measured by varying one depth cue while making the other two dimensions non-informative with respect to the depicted depth. This was achieved by fixing them to 0 (i.e. flatness). It remains questionable whether this procedure is adequate for estimating single-cue reliabilities. To underscore the results of Experiment 1, two different single-cue conditions were realized in the current study. One condition resembled the procedure of Experiment 1 (one cue relevant) whereas a second condition newly introduced stimuli that solely contained one depth cue (one cue alone).

Interestingly, both conditions were largely comparable regarding the estimated PSEs as well as the JNDs. On the one hand, this suggests that the results of Experiment 1 are valid. On the other hand, this comparability of single-cue conditions indicates that observers were able to ignore stimulus features irrelevant for the task at hand. Thus, irrelevant cues indicating flatness did not extensively bias the depth estimate towards the perception of a flat stimulus when one relevant cue indicated a different, task-relevant depth. This is an example of cue vetoing (Bülthoff & Mallot, 1988) and a clear-cut demonstration of top-down influences on multisensory depth perception (see section 2.4.4).

This result contradicts recent findings by Hillis and colleagues (Hillis, Ernst, Banks, & Landy, 2002). By comparing sensory integration mechanisms of a within-modality (vision) and a between-modality task (vision and touch), they demonstrated that mandatory fusion occurred only when cues were integrated within one sensory system. Visual stimuli consisting of discrepant texture and disparity cues led to percepts that were indiscriminable from a consistent-cue stimulus, that is, metamers occurred. This was not true for visual-haptic displays. Thus, in intramodal cases, single-cue information was lost and the observer's responses exclusively relied on the fused percept, whereas the single-cue information was still accessible in the visual-haptic condition. The data from the single-cue conditions of the present study suggest that mandatory fusion might not take place even when combining cues within one sensory system (q.v., Hogervorst, & Brenner, 2004). Conditions with a large amount of conflict between the visual depth cues, as in the single-cue task where one cue was relevant, obviously triggered a noticeable suppression of task irrelevant information and a corresponding reliance on the cue that was relevant for the task at hand.

4.4.2 Single-cue reliabilities

In the single-cue task of Experiment 2, the texture cue tended to be less reliable than shading and motion cues to depth. This result slightly differs from

Experiment 1 where the shading cue was shown to be associated with the smallest JNDs for all observers (see 3.4.1) while texture and motion cues were both substantially less reliable. What caused this discrepancy between the Experiments 1 and 2? This result can probably be attributed to the presentation time of the stimuli. In Experiment 1, each stimulus presentation lasted two seconds and consisted of one clockwise and one anticlockwise rotation for one second each. By contrast, in Experiment 2, the observer decided how long each stimulus was presented. The average presentation time in the single-cue conditions amounted to 11.255 s (*SD* = 4.315 s) in Experiment 2. Thus, on average, each observer estimated the cylinder's depth on the basis of more than five clockwise and anticlockwise movements of the texture elements. Possibly, this extension of the viewing time allowed for more precise depth estimates especially when relying on motion cues.

This hypothesis was further substantiated by calculating a 2 × 3 × 3 ANOVA on the duration of each trial using the single-cue task (one cue relevant or one cue alone), the cue (shading, texture, or motion) and the depth of the stimulus (200, 300, or 400 pixels) as within-subject factors. This analysis revealed a significant main effect of cue, $F(2, 14) = 12.78$, $\varepsilon = 1.00$, $p < .001$, $f = 0.27$. Post hoc comparisons using Tukey's HSD procedure yielded significant differences between the estimation time of the shading cue (*M* = 9.422, *SD* = 1.287) and the texture (*M* = 12.348, *SD* = 1.889) and motion cue (*M* = 11.997, *SD* = 1.468), respectively. The duration of the latter two trial types did not differ significantly. Thus, observers spent less time estimating the depth of shaded cylinders compared to stimuli whose depth was defined by texture or motions cues, respectively. As the observers freely decided when they had finished depth estimation, the results of Experiment 2 might represent maximum performance with respect to the sensitivity of each cue. Additionally, a marginally significant main effect of the method factor was obtained, $F(1, 7) = 4.48$, $p < .10$, $f = 0.15$. Overall, observers were a little slower in the single-cue condition using one relevant cue (*M* = 12.002 s, *SD* = 5.080 s) as compared to the other method using one cue alone (*M* = 10.499 s, *SD* = 3.670 s). This corresponds to the above mentioned finding of slightly lower JNDs in the latter condition (see 4.3.1), which might be attributed to a certain difficulty of suppressing task irrelevant information when the stimulus contained non-informative depth cues.

4.4.3 Dissimilarity of depth perception and slant estimation

When discussing potential limitations of Experiment 1 (see 3.4.3), the question arose whether results from slant discrimination experiments also apply to the depth estimation task used in the current studies. Specifically, slant estimation studies showed that the reliability of texture cues varies as a function of the slant of planar surfaces (Knill, 1998b; Knill & Saunders, 2003; Rosas et al., 2004). In Experiment 2, single-cue JNDs were measured separately for each depth that was used in the multiple-cue conditions, and a larger range of depths was used than in Experiment 1. It turned out that JNDs differed only slightly as a function of cylinder depth. A reanalysis of the data from Jacobs (1999) revealed a very similar pattern. In his study, reliabilities of the texture and motion cues were assessed using a comparable experimental procedure as in the single-cue

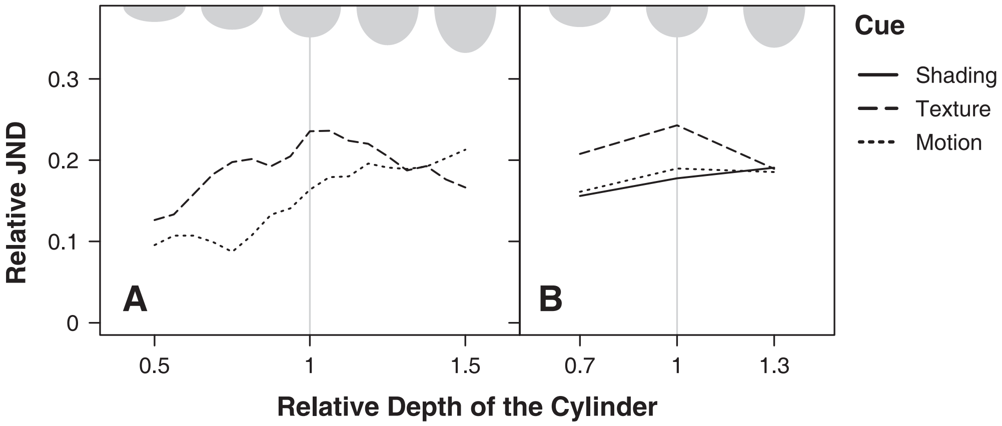


Figure 4.6. Ratio of just noticeable differences (JND) relative to the cylinder’s width as a function of the cylinder’s elongation (depth divided by width). Panel A) shows a reanalysis of the data from Jacobs (1999). Mean values across all three observers are depicted. Panel B) shows the mean values of all eight observers of Experiment 2. For illustrative purposes, cross-sections of the hemi-cylinders as a function of elongation are depicted at the top of the figure.

conditions of the current experiment where only one cue was present in the scene. Because cylinder size differed substantially between the current study and Jacobs' experiment, ratios were calculated for cylinder depth and JND, with cylinder width serving as denominator. Results are depicted in Figure 4.6. In both studies, the texture cue was less reliable when the cylinder had a circular cross section. Smaller as well as larger depths increased the reliability of this cue. By contrast, the reliability of the motion cue decreased monotonically with depth. These results do not match the predictions that would have been drawn on the basis of slant discrimination studies. In sum, depth estimation and slant estimation are not comparable and seem to require different perceptual processes.

4.4.4 Conclusions

In Experiment 2, several conditions were realized that should allow for superior predictions of the MLE model. Each stimulus was displayed as long as the observers needed for the depth estimation task and small differences between single-cue reliabilities as a function of cylinder depth were taken into account. Despite these improvements, the MLE model satisfactorily explained the empirical PSEs for less than half of the observers. Moreover, an extended Bayesian model (CMLE) that takes into account consistencies between cues outperformed the MLE model for six of eight observers. Thus, results from Experiment 1 were fully replicated. Trialwise interactions between depth cues clearly affected visual depth perception.

The CMLE model also was superior accounting for variations of the multiple-cue JNDs across observers. Whereas the MLE model consistently underestimated the JNDs, the predictions of the CLE model were in close agreement with the empirical JNDs for at least three observers. This result is especially interesting because the free parameters of the Bayesian model were only fitted to explain variations of the PSEs across conditions. The finding that these fitted parameters also account for some variation in the JNDs further substantiates the claim that an extended Bayesian model reflects the integration of visual depth estimates more adequately than does a simple MLE model, which disregards cue interactions.

5. Experiment 3

5.1 Introduction

Perhaps the main problem of studies examining sensory integration according to the MLE-framework is the estimation of single-cue reliabilities. Whereas this can be accomplished easily using several modalities (e.g., Alais & Burr, 2004; Ernst & Banks, 2002), it is difficult within one modality because stimuli have to be generated that carry only the dimension of interest. The critical assumption is that these single-cue conditions are representative for the multiple-cue condition to allow for an adequate comparison of MLE predictions and empirical data (see Knill & Saunders, 2003, p. 2555 f). Experiment 2 clearly showed that the results of different estimation methods for single-cue reliabilities were largely comparable. Thus, this aspect does not seem to be critical for Experiments 1 and 2. However, it is still possible that depth perception qualitatively differs between single-cue and multiple-cue conditions. A simple example demonstrating this possibility is shown in Figure 5.1. The image on the left side, for example, may be interpreted as depicting a slanted plane that is covered with a granite-like texture.

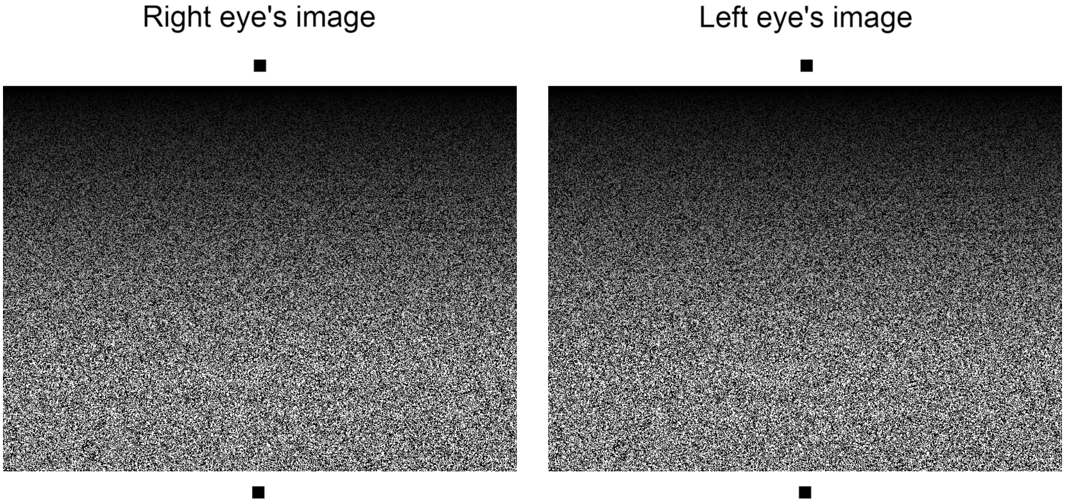


Figure 5.1. Demonstration of qualitative changes in image perception when adding disparity to a shading pattern. The stereogram can be viewed by cross-fusing both images.

However, when viewing both images using cross-fusion, a completely different percept emerges, showing vertically oriented bars in front of the slanted plane. Thus, the addition of a disparity gradient qualitatively changes the interpretation of the image. A comparable, although less obvious, alteration might have reduced the transferability of single-cue reliabilities to the multiple-cue conditions in Experiments 1 and 2. To remedy this problem, one needs a method to reliably derive single-cue weights from a multiple-cue condition. In turn, these weights could be used to predict the observer's response pattern in comparable conditions with pairwise cue consistencies and reduced single-cue reliabilities. Such a method to determine relative weights in a multiple observation task was proposed by Berg (1989). It shares several features with the perturbation analysis (Landy et al., 1995, p. 405 f.; Young et al., 1993) and it was already successfully used in the domain of psychoacoustics to examine reliability sensitive integration of tone sequences, for example (Berg, 1990).

5.1.1 Weight estimation in multiple observation tasks

The method originally described by Berg (1989) is grounded in signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 2005) and it is based on a yes-no decision task where the observer has to indicate whether a signal or noise was presented in each trial. Transferred to the current study, stimuli had to be judged whether they appeared deep or shallow. To examine the impact of several cues, deep and shallow stimuli consisted of *n* multiple cues that carried slightly different depth information x_i which were randomly sampled from one of two Gaussian distributions with different means (μ_d and μ_s) but equal variances $\sigma^2 = \sigma_d^2 = \sigma_s^2$ (see Berg, 1989, for a more general case of weight estimation that also allows for $\sigma_d^2 \neq \sigma_s^2$). It is now assumed that observers use a weighted linear combination of depth cues to judge whether the stimulus is deep or shallow. This statistic is subsequently compared to an arbitrary decision criterion *C* and the observer is assumed to answer deep (*D*) when the weighted average exceeds *C*.

The decision rule can thus be formalized to

$$\text{respond } D \text{ iff } \sum_{i=1}^{n} w_{\mathrm{i}} x_{\mathrm{i}} > C \tag{5.1}$$

This integration scheme is very similar to the MLE rule (see Ernst & Bülthoff, 2004). However, it makes no assumptions about the size of single-cue weights w_{i}. Because only relative weights are considered, it is convenient to assume that

$$\sum_{i=1}^{n} w_{\mathrm{i}} = 1 \tag{5.2}$$

Isolating x_{i} in equation 5.1 yields

$$x_{\mathrm{i}} > \left(C - \sum_{j \neq i}^{n} w_{\mathrm{j}} x_{\mathrm{j}} \right) \Big/ w_{\mathrm{i}} \tag{5.3}$$

For the sake of simplification, a new random variable Y_{i} is defined that is equal to the right side of equation 5.3. Because all x_{j} are mutually independent, and normally distributed random variables with equal variances σ^2, Y_{i} is also normally distributed with

$$\mathrm{E}(Y_{\mathrm{i}}) = \left(C - \sum_{j \neq i}^{n} w_{\mathrm{j}} E(x_{\mathrm{j}}) \right) \Big/ w_{\mathrm{i}} \tag{5.4}$$

and

$$\mathrm{Var}(Y_{\mathrm{i}}) = \sum_{j \neq i}^{n} w_{\mathrm{j}}^2 \sigma^2 \Big/ w_{\mathrm{i}}^2 \tag{5.5}$$

Whereas $\mathrm{E}(Y_{\mathrm{i}})$ differs between trials with deep and shallow stimuli, the variance $\mathrm{Var}(Y_{\mathrm{i}})$ is identical because both trial types were generated by randomly sampling from Gaussian distributions with equal variances $\sigma^2 = \sigma_{\mathrm{d}}^2 = \sigma_{\mathrm{s}}^2$. Single-cue weights w_{i} can now be calculated indirectly by using estimates of $\mathrm{Var}(Y_{\mathrm{i}})$. To this end, σ^2 is added to both sides of equation 5.5.

$$\mathrm{Var}(Y_{\mathrm{i}}) + \sigma^2 = \sum_{j \neq i}^{n} w_{\mathrm{j}}^2 \sigma^2 \Big/ w_{\mathrm{i}}^2 + \sigma^2 \tag{5.6}$$

Multiplying by w_{i}^2 yields

$$w_{\mathrm{i}}^2\left(\mathrm{Var}(Y_{\mathrm{i}}) + \sigma^2\right) = \sum_{j=1}^{n} w_{\mathrm{j}}^2\sigma^2 \tag{5.7}$$

Thus, for all cues *i*, the right side of equation 5.7 is identical and for any arbitrary cue combination *j* and *k*, one can state that

$$w_{\mathrm{j}}^2\left(\mathrm{Var}(Y_{\mathrm{j}}) + \sigma^2\right) = w_{\mathrm{k}}^2\left(\mathrm{Var}(Y_{\mathrm{k}}) + \sigma^2\right) \tag{5.8}$$

Isolating both weights on one side of the equation yields

$$\frac{w_{\mathrm{j}}^2}{w_{\mathrm{k}}^2} = \frac{\left(\mathrm{Var}(Y_{\mathrm{k}}) + \sigma^2\right)}{\left(\mathrm{Var}(Y_{\mathrm{j}}) + \sigma^2\right)} \tag{5.9}$$

The variance of single cues σ^2 is known and Var(Y_{i}) can be estimated from the empirical data by using conditional on single stimulus (COSS) functions (Berg, 1989, p. 1744). These functions are generated for deep and shallow stimuli separately by plotting the proportion of "deep" responses as a function of each single-cue depth x_{i} irrespective of the remaining cues. By fitting cumulative Gaussians to these data, Var(Y_{i} | deep) and Var(Y_{i} | shallow) can be calculated (see Figure 5.2 for an illustration). Since COSS functions for deep and shallow stimuli theoretically have the same slope for any given *i* (see equation 5.5), the mean of the two estimates is used as an estimate of Var(Y_{i}). By inserting Var(Y_{i}) and σ^2 into equation 5.9, single-cue weights can be determined iteratively. As a final step, the weights are normalized to satisfy assumption 5.2.

These calculations are illustrated in Figure 5.2 for a set of simulated data. The following algorithm was used: First, depth values for three cues (shading, texture, and motion) were randomly sampled for 1000 trials from either a deep (μ_{d} = 90 cm) or a shallow distribution (μ_{s} = 70 cm) with equal variances $\sigma_{\mathrm{d}}^2 = \sigma_{\mathrm{s}}^2 = 400\ \mathrm{cm}^2$. Afterwards, a combined depth estimate was generated by computing the weighted average of single cues using the following weights: $w_{\mathrm{S}} = .42$, $w_{\mathrm{T}} = .23$, $w_{\mathrm{M}} = .35$. Gaussian noise (μ = 0 cm, $\sigma^2 = 9\ \mathrm{cm}^2$) was added and the response pattern was calculated for a decision criterion of C = 80 cm (see equation 5.1). Using these data, COSS functions were generated for deep and shallow trials separately according to the procedure described above. As can be seen from Figure 5.2, COSS functions for deep and shallow trials are parallel for all cues and the function's slope indicates the perceptual weight of the respective cue.

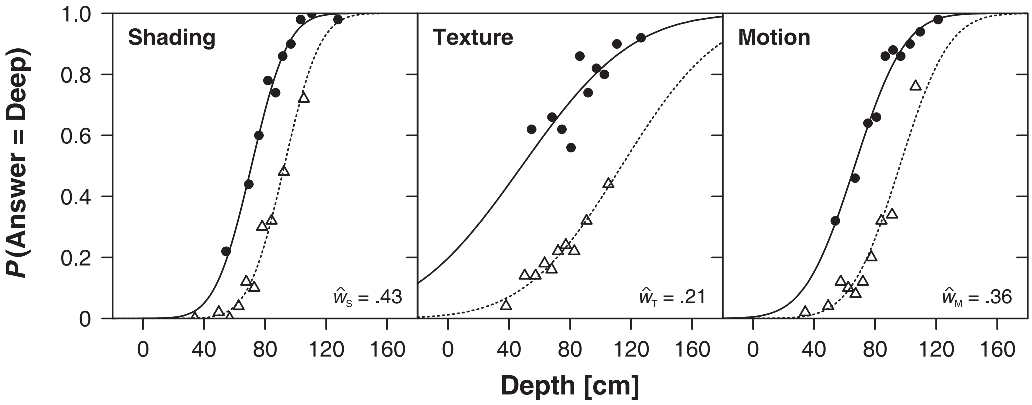


Figure 5.2. Conditional on single stimulus (COSS) functions for a set of simulated data (see main text). Circles and solid lines depict the response pattern for deep stimuli; triangles and dotted lines show the corresponding pattern for shallow stimuli. The estimated weights for the current data that closely resemble the weights that were used for the simulation are printed in the lower right corner of each panel.

5.1.2 The multiple observation task in visual depth perception

The logic of the multiple observation task can be easily transferred to studies on sensory integration in visual depth perception. This can be accomplished by generating two classes of stimuli (deep and shallow) with mutually independent single-cue depths. The observer's response pattern serves to derive single-cue weights and these values can subsequently be used to predict the response pattern for other classes of stimuli. It is the main advantage of this method that potentially problematic single-cue stimuli are not required. Thus, predictions are based on the results of a multiple-cue condition instead of several single-cue tasks. This procedure was used for the current and the following experiment. The current study aimed at providing a first test of the multiple observation task in the domain of visual depth perception. Experiment 4 used this task to examine whether the embedded additive integration scheme is sufficient to predict the empirical data under conditions of pairwise cue consistencies and reduced single-cue reliabilities, respectively.

5.2 Method

5.2.1 Stimuli and apparatus

Hemi-cylinders with an elliptical cross section served as stimuli. Instead of viewing them on a small two-dimensional video display as in Experiment 1 and 2, they were displayed on a large rear projection screen (260 × 192.5 cm) with a color depth of 32 bits. The projection allowed for stereoscopic viewing by use of two projectors with a resolution of 1400 x 1050 pixels each. The light of the two projectors was linearly polarized in orthogonal planes. Participants wore matching polarization filters such that each eye received a unique image. The background of the projection screen was uniformly colored in dark blue to attenuate the occurrence of ghosting due to an imperfect separation of both visual channels. The observer was seated 300 cm in front of the screen with her eye height aligned to the center of the projection screen. One semiaxis of the elliptical cross section which lay in the projection plane was held constant (80 cm) and represented the cylinder's width. The other semiaxis that was parallel to the observer's line of sight corresponded to the cylinder's depth and was varied across trials.

The depth of the cylinder was defined by four visual cues: 1) A homogeneous texture consisting of white circles was mapped on the cylinder's surface. Between 90 and 120 circles per square meter were randomly placed on a two-dimensional sheet that was sized 300 × 250 cm. The radius of each circle was randomly sampled from a uniform distribution ranging from 0.5 to 4.5 cm. The distance between the boundaries of any two circles was at least 2 cm. The final texture was mapped on the surface of a cylinder. In contrast to Experiments 1 and 2, the space between the circular texture elements was transparent, that is, the background was visible between them (see Figure 5.3, panel A). Cylinder depth could be derived from variations in size, shape, and density of texture elements in certain parts of the display (Blake et al., 1993; Buckley et al., 1996; Stevens, 1981; Todd & Akerstrom, 1987). 2) The texture elements had Lambertian reflectance properties and were illuminated by a laminar light source from the direction of the observer causing a shading gradient that was visible within each circle and across texture elements. Because of the otherwise transparent texture, the background colour was visible between the shaded circles (see Figure 5.3, panel B). Thus, shading information was present within the texture elements itself instead of the cylinder's background (cf., Experiments 1 and 2). This modification was implemented to prevent observers from switching their focus between background and foreground to estimate the cylinder's depth. As in the Experiments 1 and 2, however, the gradient was normalized to obscure the observer-object-distance. The gradient's shape provides depth information of the displayed hemicylinder (Mingolla & Todd, 1986; Pentland, 1989). Because this gradient was distributed across several texture elements, it might be supposed that depth from shading judgements become less reliable in this study as compared to Experiment 1 and 2. 3) The shaded and textured hemi-cylinder was stereoscopically presented on the above mentioned rear projection screen by taking into account individual eye bases (see Figure 5.3, panel C). The resulting disparity gradient provides relative depth information and – when scaled by distance – also absolute depth (Johnston, 1991; Johnston et al., 1994; Rogers & Bradshaw, 1995). 4) Using a comparable procedure as in Experiments 1 and 2, a motion cue was constructed by moving texture elements horizontally along the surface of the cylinder with a constant velocity either in clockwise or anticlockwise direction. This was achieved by computing the position of the centre of each texture element as a function of time when travelling on a track defined by the elliptical cross section of the visible hemi-cylinder. The velocity was drawn from a uniform distribution that ranged from 15 to 25 cm per second. It was always constant for each texture element within one trial but varied between trials to prevent depth judgements based on absolute speed and maximum displacement of texture elements within the scene (Johnston et al., 1994). The cylinder's depth could be estimated using velocity and acceleration gradients that were produced by the motion cue (Perotti et al., 1998).

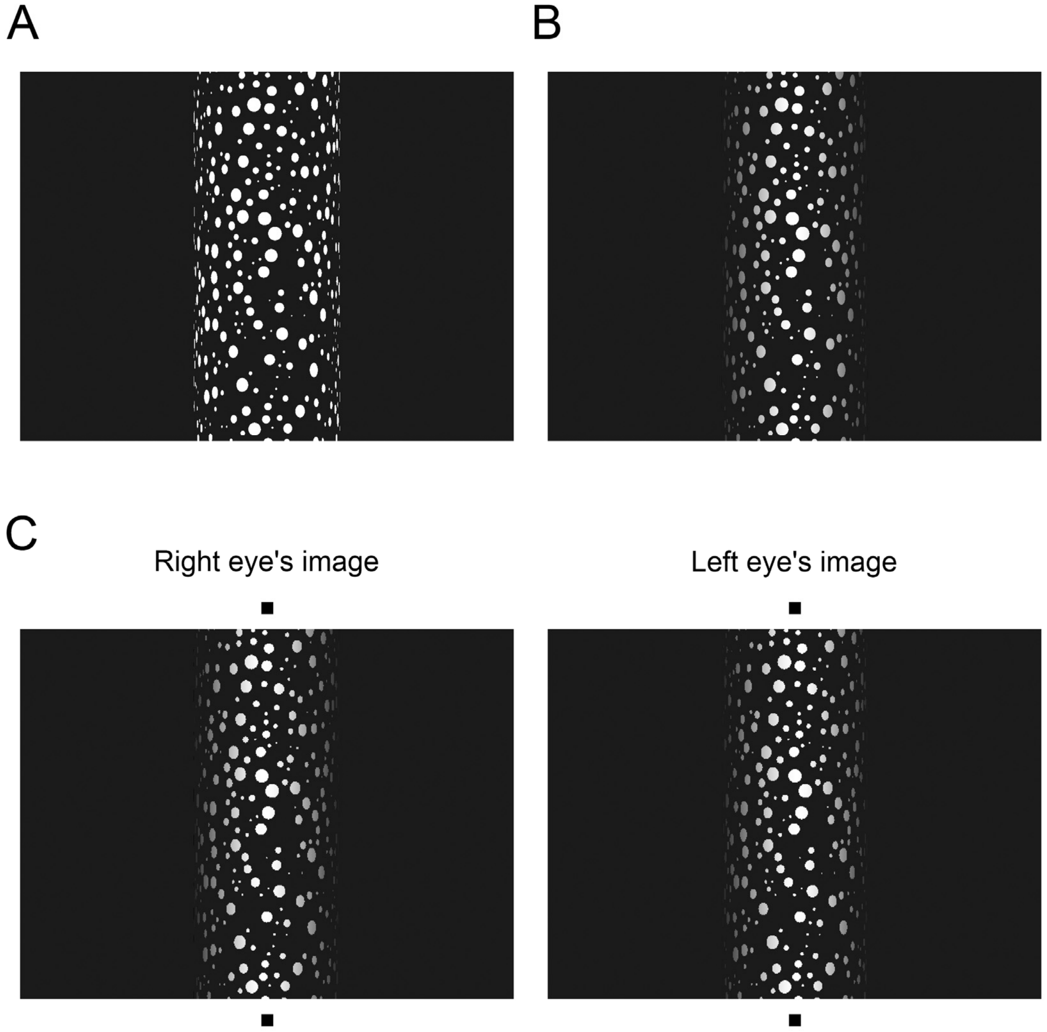


Figure 5.3. Illustration of the stimulus generation algorithm: In a first step, a homogeneous texture consisting of white circles was mapped on the cylinder's surface (panel A). Afterwards, the texture elements were illuminated by a laminar light source which caused a shading gradient (panel B). This stimulus was stereoscopically presented (panel C; stereogram can be viewed by cross-fusing both images) and a motion cue was generated (see text). The background that was also visible between texture elements was uniformly coloured in dark blue. The stimulus shown here has consistent cues and its depth to width ratio is larger than one.

All four depth cues (texture, shading, stereo, and motion) could be manipulated orthogonally. Thus, arbitrary combinations could be created. This technique allowed for an examination of the impact of single cues on the perceived depth of the cylinder using a multiple observation task (see 5.1.1; Berg, 1989, 1990; Lutfi, 1995). Stimuli were generated in real time using OpenGL on a Pentium 4 3.0 GHz computer with a NVIDIA Quadro FX 3000 graphics chip. They were stereoscopically displayed on the above mentioned rear projection screen with a color depth of 32 bits. The video refresh rate was set to 60 Hz. The cylinder's width was set to 80 cm and its height equaled the vertical screen resolution (i.e. 192.5 cm), thus, the cylinder's bottom and top were never visible to the observer. For above mentioned reasons, the background was uniformly colored in dark blue with a luminance of 0.16 cd/m^2. The stimulus was displayed in grayscales and the luminance in the centre of the shading gradient was 21.91 cd/m^2. The ceiling lighting of the laboratory was switched off during the course of the experiment. The stimuli were viewed binocularly from a distance of 300 cm and their image subtended 15.2° of visual angle in the horizontal and 65.4° in the vertical dimension.

5.2.2 Experimental design

The present experiment was carried out to answer two research questions: First, it should be tested whether the multiple observation task that was already successfully used in the domain of psychoacoustics (Berg, 1989, 1990), can be reasonably implemented in the area of visual depth perception. Second, it should be determined how much trials are necessary to obtain stable estimates of single-cue weights and observer sensitivity. To this end, participants had to judge the depth of hemi-cylinders that were composed of the four above mentioned depth cues that were varied independently from each other and analysed according to the procedure described in 5.1.1.

5.2.3 Procedure

Stimulus generation and presentation: Test stimuli were generated by independently sampling one depth for each of the four cues from either a Gaussian distribution with an expected value of μ_d = 90 cm (deep stimulus) or from a distribution with μ_s = 70 cm (shallow stimulus). Both distributions had a standard deviation of σ = 20 cm. Using this procedure, 500 stimuli were constructed for each distribution and presented on the projection screen with one clockwise and one anticlockwise rotation that lasted for one second each. The 1000 test trials were randomly allocated to three experimental sessions with 300 trials in the first and 350 trials in the second and third session, respectively. No feedback was provided during testing.

To become familiar with the procedure and to establish a decision criterion, a training session was performed in advance that was not considered in the statistical analyses. This session consisted of 200 "easy" trials with cue depths sampled from either a Gaussian distribution with an expected value of μ_d = 100 cm (σ = 20 cm; deep stimulus) or μ_s = 60 cm (σ = 20 cm; shallow stimulus). An additional amount of 200 "difficult" training trials was created using the properties of the test stimuli (see above). Feedback was given after each training trial. Test sessions also started with a training period consisting of 100 "easy" and 100 "difficult" trials (test session 1) or 50 "easy" and 100 "difficult" trials (test session 2 and 3), respectively. These training trails were also excluded from the statistical analyses.

Specification of the observer's task: A one interval 2AFC task was used in this experiment. In each trial, a hemi-cylinder that was constructed using the procedure described above (see 5.2.1) was displayed on the projection screen. Afterwards, two gray buttons labeled deep (1) and shallow (2) appeared on the screen and the observer had to judge whether the stimulus appeared to be deep or shallow by pressing the corresponding key on the computer keyboard. In the training trials, the color of the activated button changed to green or red depending on whether the response was correct or wrong, respectively. Additionally, the words “correct” or “wrong” were displayed at the top of the screen. In the test trials, the button’s color changed to light gray to indicate which button was pressed. The next trial started one second after the observer’s response. No time limit was specified for the task but the response time was saved for each trial. After each block consisting of 50 trials, a cumulative feedback of the proportion of correct responses was given. The observer could freely decide when to start the next block by pressing the “Enter” key.

Observers participated in the experiment for 4 sessions (one training and three test sessions) that usually occurred on different days. Each session lasted no more than 45 minutes and all sessions were accomplished within a period of 2 to 18 days (*M* = 8 days; *Mdn* = 8 days). After the experiment, participants completed a questionnaire (see Appendix D) that consisted of several questions asking for strategies and cues that were used to accomplish the task.

5.2.4 Parameter extraction and statistical analyses

Using the procedure that is described in 5.1.1, single-cue weights were obtained. Moreover, *d’* was calculated as an overall index of the observer’s sensitivity according to the following equation (see Macmillan & Creelman, 2005, p. 8):

$$d' = \Phi^{-1}(H) - \Phi^{-1}(FA) \tag{5.10}$$

Here, $\Phi^{-1}(p)$ is the p-th quantile of the standard normal distribution, *H* and *FA* are the proportions of hits (correctly classified deep stimuli) and false alarms (incorrectly classified shallow stimuli), respectively. Additionally, response times were defined as the interval from the stimulus offset to the observer’s key press. Response times larger than 3 s, which occurred in only 0.03% of all trials on average, were excluded. All these dependent variables were separately calculated for the first (trials 1 to 500) and second half of the experiment (trials 501 to 1000) as well as for the whole study.

All statistical analyses were carried out using R (version 2.4.1) or SPSS (version 14.0.2). An a priori alpha level of .05 was used for all statistical tests. To correct for potential violations of the sphericity assumption in repeated-measures ANOVAs involving more than one degree of freedom in the enumerator, the Huynh-Feldt procedure (Huynh & Feldt, 1976) was applied. Cohen's *f* is reported as an effect size estimate for each statistically significant effect (Cohen, 1988, p. 273 ff.).

5.2.5 Participants

Eight women who had not participated in Experiment 1 or 2 took part in this study. They were psychology students or postgraduates and their age ranged from 19 to 32 years (M = 25 years). All gave written informed consent before the experiment started. They had normal or corrected-to-normal vision and were naive to the purposes of the experiment.

During the course of the study, no participant spontaneously reported having noticed discrepancies between depth cues. In the questionnaire that was completed after the experiment (see Appendix D), four observers noted that they were incidentally surprised about the feedback in the training period because they were quite sure that the respective stimulus belonged to the other depth category. This might be due to conflicting depth cues but may also be related to the large overlap between the distributions of deep and shallow stimuli. To test whether both these groups of participants differed in their response pattern, all statistical analyses were also carried out using this variable as between-subject factor.

5.3 Results

In a first step, a 2 × 4 ANOVA was conducted on the single-cue weights using the within-subject factors experimental block (trials 1-500 vs. 501-1000) and cue (shading, texture, motion, disparity). Only a significant main effect of cue was obtained, $F(3, 21) = 17.88$, $\varepsilon = .78$, $p < .001$, $f = 1.53$. The interaction of experimental block and cue was clearly insignificant, $F(3, 21) = 1.18$, $\varepsilon = .85$, *p* = .34, *f* = 0.06, thus, the estimated weights were largely comparable between the first and second half of the experiment (see Figure 5.4). Disparity and motion were weighted higher than shading and texture cues, but a series of *t*-tests for significant differences from 0 revealed that each cue received a positive weight across all trials of the experiment: shading, *t*(7) = 3.72, *p* < .01; texture, *t*(7) = 5.19, *p* < .01; motion, *t*(7) = 6.96, *p* < .001; disparity, *t*(7) = 9.62, *p* < .001.

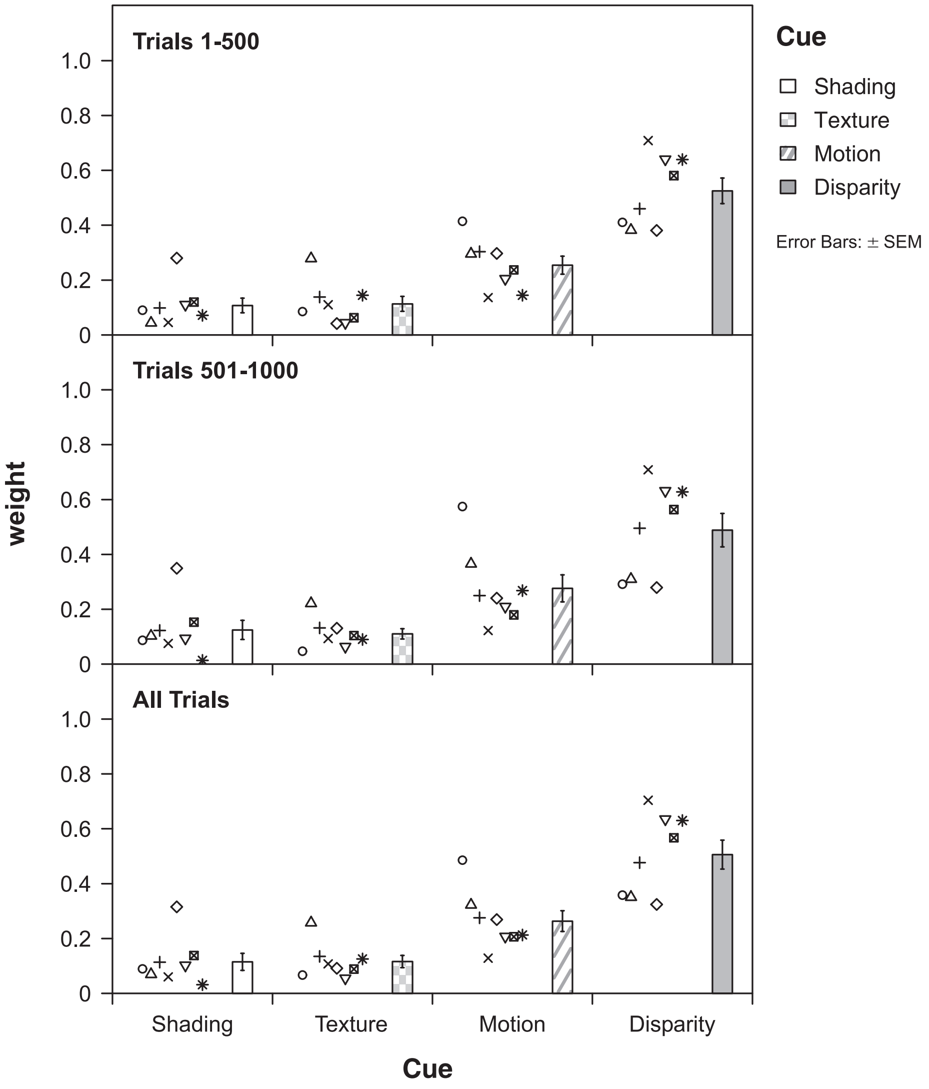


Figure 5.4. Perceptual weights for each cue depicted separately for the first (trials 1-500) and second half of the experiment (trials 501-1000) as well as for all trials of the current study. Different symbols show the data of each observer; bars correspond to mean values across participants.

An extension of the above mentioned ANOVA by a between-subject factor reflecting the potential detection of cue conflicts, led to fully comparable results. All interactions with the between-subject factor were statistically insignificant: conflict awareness × cue, $F(3, 18) < 1, f = 0.11$; conflict awareness × experimental block × cue, $F(3, 18) < 1, f = 0.02$. Thus, participants did not change their cue weighting when becoming aware of conflicting depth information.

The *d'* values were also comparable between the first and second block of the experiment and did not differ significantly as revealed by a paired *t*-test, $t(7) = 1.03$, $p = .34$, $f = 0.04$ (see Figure 5.5). To check whether a potential detection of cue conflicts affected the observers' sensitivity, a 2 × 2 ANOVA was conducted on the *d'* values using conflict awareness as between-subject factor and experimental block as within-subject factor. As expected, the main effect of experimental block was statistically insignificant, $F(1, 6) = 1.16, p = .32, f = 0.04$. Furthermore, the interaction of conflict awareness and experimental block, $F(1, 6) = 1.68, p = .24, f = 0.05$, as well as the main effect of experimental block did not reach statistical significance, $F(1, 6) < 1, f = 0.32$. Thus, *d'* did not differ between observers that were potentially aware of cue conflicts and participants that did not notice such conflicting depth information.

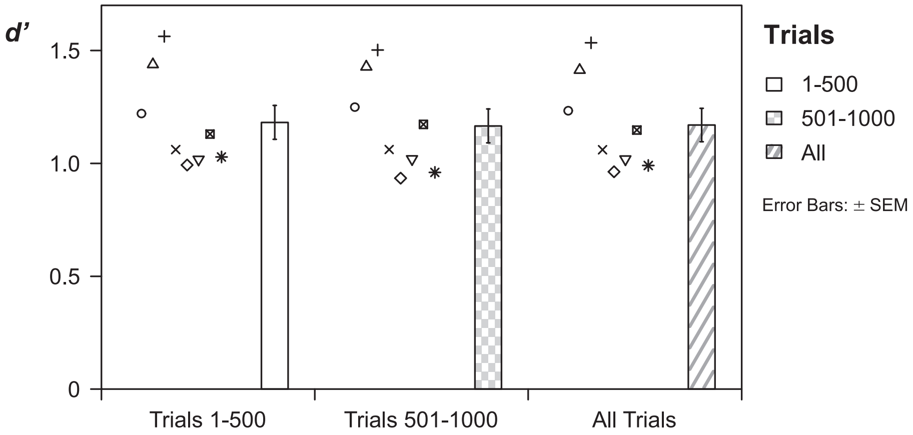


Figure 5.5. Observer sensitivity as indicated by *d'* for the first (trials 1-500) and second half of the experiment (trials 501-1000) as well as for all trials of the current study. Different symbols depict the data of each observer; bars correspond to mean values across participants.

In a final step, response times were compared between the first and second half of the experiment. They did not differ significantly as revealed by a paired *t*-test, $t(7) < 1, f = 0.03$, and amounted to 254.94 ms on average ($SD = 91.99$ ms). Overall, responses were given very fast. Thus, observers quickly decided whether a given stimulus appeared deep or shallow and pressed the corresponding key briefly after stimulus offset. Additionally, a 2 × 2 ANOVA was conducted on the response times using conflict awareness as between-subject factor and experimental block as within-subject factor. The main effects of experimental block, $F(1, 6) < 1, f = 0.03$, and conflict awareness did not reach statistical significance, $F(1, 6) < 1, f = 0.15$. However, a marginally significant interaction of conflict awareness and experimental block was obtained, $F(1, 6) = 4.18, p < .10$, $f = 0.18$. Participants who potentially detected conflicting depth information had slightly smaller response times in the first experimental block as compared to the other group of observers. This difference disappeared in the second half of the experiment (see Table 5.1).

Table 5.1. Response times as a function of experimental block for observers who potentially became aware of cue conflicts and participants who did not notice conflicting depth information. Additionally, values were averaged across all trials of the current study.

	Trials 1-500		Trials 501-1000		All Trials	
	M	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Aware	228.69	(75.59)	255.27	(92.54)	241.98	(83.01)
Unaware	286.49	(99.15)	249.29	(128.19)	267.89	(111.39)

5.4 Discussion

Overall, the largest perceptual weight was assigned to the disparity cue. Shading and texture affected the perceived depth to a much smaller degree. However, all four cues were taken into account for judging the hemicylinders' depth. This is especially interesting because in the post-experimental questionnaire, only one observer reported that she based her decision on all four cues. Six out of eight observers failed to notice that variations in the shading pattern could be used to infer the cylinders' depth. Thus, the responses that were given in the perceptual task did not depend on a conscious representation of depth cue variations. Observers performed the task more or less intuitively by taking into account all cues that provided information about the depth of the current stimulus.

The comparably large influence of the disparity cue on the perceived depth of the hemicylinders is consistent with previous research using homogeneous textures (Johnston et al., 1993) or a pattern of randomly distributed lines (Adams & Mamassian, 2004). Buckley and Frisby (1993) reported comparable results for the depth perception of real models of textured ridges (Experiment 5) and Ernst et al. (2000) found a relatively large influence of the stereo cue on slant perception even when compared to the weight of a highly regular texture that did also include perspective cues. Tittle and colleagues (1998) reported a different pattern of stereo, texture, and shading weights for shape and curvature judgments, but the disparity cue received the largest weight in both tasks which is also comparable to the results of the present experiment.

Interestingly, the shading cue affected the perceived depth of the hemicylinder to a much smaller degree than in Experiment 1. This might be related to the stimulus generation algorithm that changed in this study. Here, the texture elements themselves instead of the space between them were shaded to prevent observers from switching their focus between background and foreground. Thus, the shading gradient was distributed across several texture elements. Obviously, this modification impeded global depth judgements and led to a downweighting of this cue (see Koenderink & van Doorn, 1995, for comparable results on the perception of pictorial relief using spatially separated probe points).

Altogether, the multiple observation task that was implemented in the current experiment seemed to be useful for deriving perceptual weights from a multiple-cue condition. Observers took into account all available cues when judging the hemicylinders' depth and showed a very stable response pattern. The first 500 trials were highly representative for the whole experiment with respect to the perceptual weights as well as to the sensitivity index *d'*. On the basis of these results, Experiment 4 was designed to test whether the MLE model sufficiently explains the integration of visual depth cues in a multiple observation task.

6. Experiment 4

6.1 Introduction

It was demonstrated in Experiment 3 that the multiple observation task (see 5.1; Berg, 1989, 1990; Lutfi, 1995) can be successfully used to derive single-cue weights for different visual depth cues. In the current Experiment, the same task was used to test the predictions of a simple additive cue integration scheme. Therefore, three visual cues (shading, texture, and disparity) were selected to define the depth of visual hemi-cylinders. Only static displays were used, thus, the motion cue was eliminated from the stimulus. This modification was implemented because of three reasons: First, in the previous Experiments, the stimulus did not rotate itself; instead, the texture moved along its surface. Although some real-world stimuli comprise comparable motion cues (e.g., a rotating advertising pillar), this kind of moving texture can be assumed to have low ecological validity because it can be rarely seen outside the laboratory. Second, the texture cue may be confounded to the motion cue to an unknown degree because visual motion emerges from moving texture elements and motion cannot be visualized without texture (see footnote 4, Experiment 2). A recent study by Rosas and colleagues (2007) also suggests that combined texture and motion cues might be processed cooperatively in a non-linear manner suggesting some form of strong fusion. In the current study, this potential confound should be avoided. Third, using different depth cues that can be varied independently from each other offers the possibility to examine the generalizability of the impact of cue interactions in the process of depth cue integration.

In the current Experiment, several conditions were realized to test the adequacy of an additive integration model that is sensitive to single-cue reliabilities. First, a control condition similar to Experiment 3 was generated to allow for the calculation of single-cue weights and *d'*. Second, conditions with pairwise cue consistencies were created. For these conditions, an additive linear integration rule predicts that cue pairings should be weighted according to the sum of both single-cue weights. Moreover, the observer's sensitivity *d'* is assumed to decline because only two instead of three independent estimates are available to judge the cylinder's depth. In a third set of conditions, single-cue reliabilities were reduced and it was predicted that this would lead to a reduced weight for the corrupted cue and a smaller *d'* because of the increased variability of the combined estimate. These qualitative predictions are tested in section 6.3.1. Furthermore, quantitative predictions of the MLE model, which is only a special case of an additive integration scheme that is sensitive to single-cue reliabilities, are derived and tested in section 6.3.2. Because Experiment 3 revealed sufficiently stable parameter estimates when using the first half of all trials, only 500 trials for each condition were used in this study to reduce total experimental time.

6.2 Method

6.2.1 Stimuli and apparatus

As in the previous Experiments, hemi-cylinders with an elliptical cross section served as stimuli. They were generated using the same algorithms as in Experiment 3 with two exceptions: First, only static displays were used, thus, the motion cue was eliminated from the stimulus for reasons described above. Second, the stimulus generation algorithm allowed for a manipulation of single-cue reliabilities in this experiment. This was achieved by assigning an array of three depth values for the respective cues to each texture element. Thus, the cylinder was constructed of texture elements that could stem from stimuli with different depths even within one cue dimension. Increasing the variability of depths across texture elements for one cue should reduce its reliability. A comparable although less precisely adjustable technique was used by Young et al. (1993) who reduced the reliability of a texture cue by using ellipses instead of circular texture elements. In the current experiment, the variability of depths within one cue was realized by randomly sampling depths for each texture element from a Gaussian distribution with μ = 0 cm and σ = 50 cm. These values were added to the overall depth that was assigned to this cue. Thus, single-cue reliabilities could be manipulated independently from single-cue depths of all cues. The stimulus generation algorithm prevented the occurrence of negative depths by repeating the random sampling whenever a negative value was drawn. Example stimuli with reduced reliability of shading, texture and disparity are depicted in Figure 6.1.

The stimuli were displayed for 1 s each on the same large rear projection screen as in Experiment 3 (260 × 192.5 cm) using the same resolution (1400 × 1050 pixels, 60 Hz) and color depth (32 bits). The observer was seated 300 cm in front of the screen with her eye height aligned to the center of the projection screen. One semiaxis of the cylinder's elliptical cross section which lay in the projection plane was held constant (width = 80 cm). The other semiaxis that was parallel to the observer's line of sight corresponded to the cylinder's depth and was varied across trials. The three depth cues shading, texture, and disparity could be manipulated independently of each other with respect to mean depth and reliability. The cylinder's height equaled the vertical screen resolution (i.e., 192.5 cm). Thus, the cylinder's bottom and top were never visible to the observer. All other details regarding the stimulus generation and presentation were identical to Experiment 3.

6.2.2 Experimental design

The Experiment was conducted to test whether a simple additive combination rule can explain the integration of visual depth cues in a multiple observation task. Therefore, three different conditions were realized: 1) In a control condition, the three depth cues shading, texture, and disparity were independently varied to estimate single-cue weights and observer sensitivity *d'* according to the logic of the multiple observation task (see 5.1; Berg, 1989, 1990; Lutfi, 1995). 2) These values were subsequently used to predict weights and *d'* for different pairs of cue consistencies. 3) Additionally, the impact of lowering the reliability of single cues

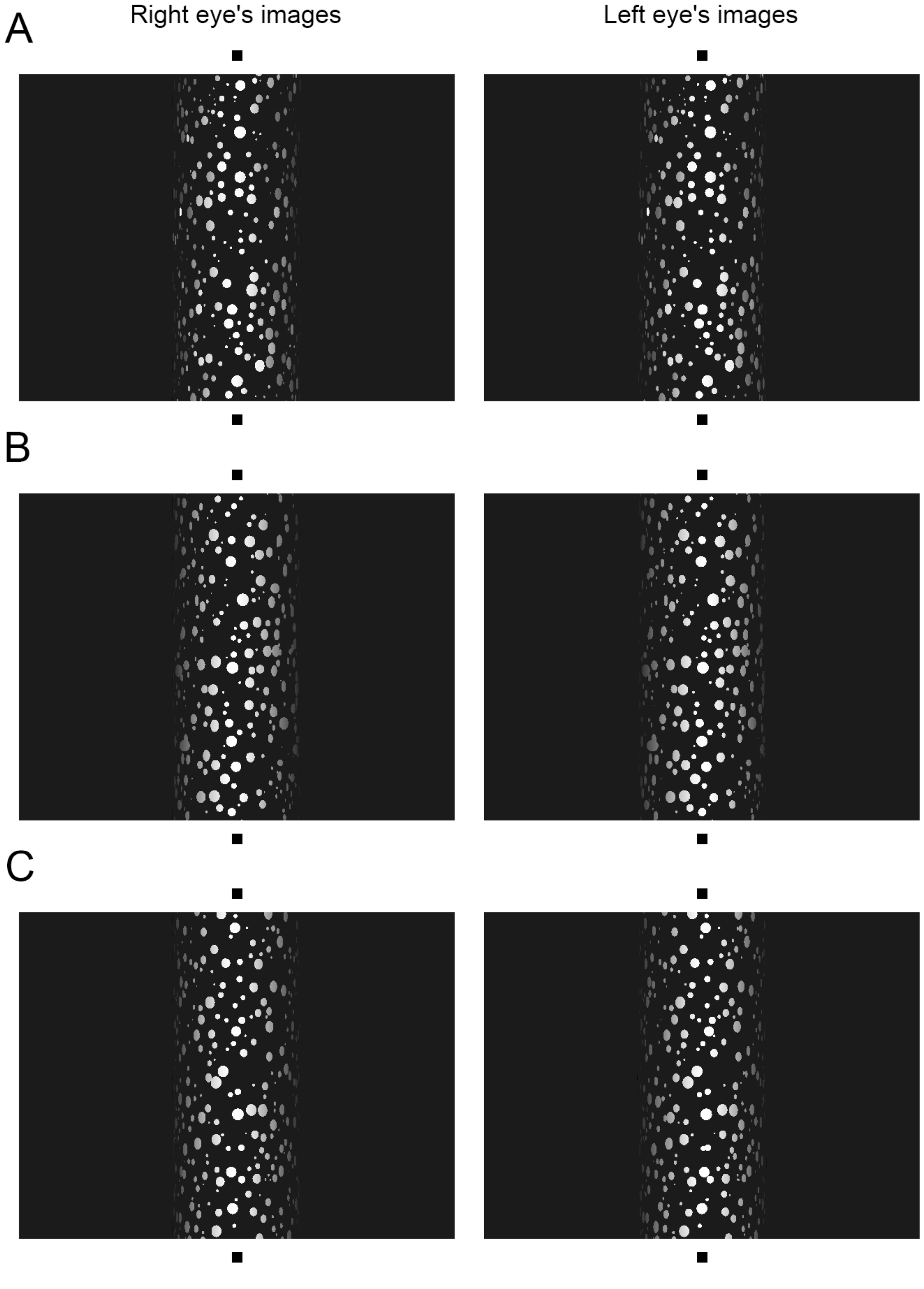


Figure 6.1. Examples of a cylinder with a relatively less reliable shading (panel A), texture (panel B), and disparity cue (panel C), respectively. The reliability of both other cues is not reduced. Stereograms can be viewed by cross-fusing both images. In the Experiment, the background was uniformly coloured in dark blue. The mean depth of all cues is consistent with a depth to width ratio larger than one for all stimuli shown here.

on their weighting and the observer's sensitivity was estimated in a third condition. These values were also tested against the predictions of the linear cue combination scheme.

6.2.3 Procedure

Stimulus generation and presentation: Test stimuli were generated using the same properties as in Experiment 3. In the control condition, the depth of each cue was determined by independently sampling from either a Gaussian distribution with an expected value of μ_d = 90 cm (deep stimulus) or from a distribution with μ_s = 70 cm (shallow stimulus). Both distributions had a standard deviation of σ = 20 cm. Single-cue reliabilities were not reduced in this condition and 250 stimuli were constructed for each distribution. The same amount of stimuli was generated with pairwise consistencies between shading and texture, shading and disparity, and texture and disparity cues by assigning the same depth to the consistent cues while independently determining the depth of the deviant cue according to the Gaussian sampling described above. In this condition, all single-cue reliabilities remained unchanged, too. In a third condition, single-cue reliabilities were reduced while allowing for discrepant depths across cues. The reliability of the shading, texture, or disparity cue, respectively, was lowered by introducing variability of depths across texture elements within this cue. The depths were varied according to a Gaussian distribution with μ = 0 cm and σ = 50 cm. All conditions are specified in Table 6.1. Altogether, 3500 test trials were constructed and randomly allocated to nine experimental sessions with 400 trials in the first to eighth and 300 trials in the ninth session, respectively.

As in Experiment 3, a training session was performed in advance. This session consisted of 250 "easy" trials with cue depths sampled from either a Gaussian distribution with an expected value of μ_d = 100 cm (σ = 20 cm; deep stimulus) or μ_s = 60 cm (σ = 20 cm; shallow stimulus) and 250 "difficult" trials that were sampled from the same distributions as the test stimuli (see above). Test sessions also started with a training period consisting of 50 "easy" and 50 "difficult" trials. For the training trials, cue depths were sampled independently and single-cues reliabilities remained unaltered. Thus, "difficult" training trials corresponded to trials of the control condition. The other conditions were not practiced to prevent observers from adjusting their criterion to condition specific features instead of developing a general cue combination rule. Feedback was given after each training trial and these trials were not included in the analyses.

Table 6.1. Parameters of the Gaussian sampling procedure that was used to determine single-cue depths and reliabilities for the different experimental conditions.

		Depth μ (σ = 20 cm)			Reliability σ (μ = 0 cm)		
Condition	n_{Trials}	μ_S	μ_T	μ_D	σ_S	σ_T	σ_D
Control	250	70	70	70	0	0	0
	250	90	90	90	0	0	0
Pairwise consistencies	250	70	$\mu_T = \mu_S$	70	0	0	0
	250	90	$\mu_T = \mu_S$	90	0	0	0
	250	70	70	$\mu_D = \mu_S$	0	0	0
	250	90	90	$\mu_D = \mu_S$	0	0	0
	250	70	70	$\mu_D = \mu_T$	0	0	0
	250	90	90	$\mu_D = \mu_T$	0	0	0
Reduced Reliability	250	70	70	70	50	0	0
	250	90	90	90	50	0	0
	250	70	70	70	0	50	0
	250	90	90	90	0	50	0
	250	70	70	70	0	0	50
	250	90	90	90	0	0	50
Total	3500						

Note. The characters S, T, and D denote the shading, texture and disparity cue, respectively. The unit of all values except the number of trials is cm.

Specification of the observer's task: A one interval 2AFC task was used. The stimulus was displayed for 1 s and the observer had to indicate whether it appeared deep or shallow by pressing the corresponding key on a computer keyboard. The button press and the feedback for the training trials was visualized as in Experiment 3 (see section 5.2.3) The next trial started one second after the observer's response. No time limit was specified for the task but the response time was saved for each trial. Cumulative feedback of the proportion of correct responses was given after each block consisting of 50 test trials to maintain the observer's attention and motivation. The observer could freely decide when to start the next block by pressing the "Enter" key.

The observers participated in ten experimental sessions that lasted about 35 to 45 minutes each and usually occurred on different days. All sessions were accomplished within 7 to 27 days (*M* = 13 days; *Mdn* = 11 days). After the last experimental session, participants completed a questionnaire (see Appendix E) that consisted of several questions asking for strategies and cues that were used to accomplish the task. All observers indicated that they used shading, texture, and disparity cues to estimate the cylinder's depth. Moreover, examples of stimuli from each condition were randomly generated and displayed on the projection screen. The observer had to judge the difficulty of each condition on a 4-point Likert scale ranging from high to low. Moreover, potential changes in strategies could be freely described for the respective condition in the questionnaire.

6.2.4 Parameter extraction and statistical analyses

Single-cue weights and observer sensitivity *d'* were calculated for each condition according to the logic of the multiple observation task (see 5.1). In the conditions with pairwise cue consistencies, one weight was estimated for both consistent cues and one for the deviant cue, respectively. Specific weights for each cue could be determined for the remaining conditions, because cue depths were independently drawn from a Gaussian distribution in these cases. After qualitatively evaluating the adequacy of a simple linear integration model that is sensitive to cue reliabilities in section 6.3.1, an MLE model is outlined in section 6.3.2 and weights and *d'* are quantitatively compared to the empirical data.

All statistical analyses were carried out using R (version 2.4.1) or SPSS (version 14.0.2). An a priori alpha level of .05 was used for all statistical tests and the Huynh-Feldt procedure (Huynh & Feldt, 1976) was applied to correct for potential violations of the sphericity assumption in repeated-measures ANOVAs involving more than one degree of freedom in the enumerator. Cohen's *f* is reported as an effect size estimate for each statistically significant effect in the ANOVAs (Cohen, 1988, p. 273 ff.).

6.2.5 Participants

Ten observers (4 women, 6 men), naïve to the purposes of the Experiment, participated voluntarily after giving written informed consent. Nine observers were psychology students or postgraduates and one was a student of mathematics. Their age ranged from 23 to 35 years (*M* = 27 years) and all had normal or corrected-to-normal vision. Observers BK and DK had also participated in Experiment 2. No participant spontaneously reported having noticed discrepancies between depth cues but when specifically asked after the experiment, two observers reported that they sometimes noticed conflicting depth cues. All observers became aware of the alteration of single-cue reliabilities but none consciously perceived pairwise cue consistencies. In the post-experimental questionnaire, nearly all observers (9 of 10) reported that they often judged the cylinder's depth intuitively and did not employ specific task strategies.

6.3 Results

6.3.1 Adequacy of an additive linear integration rule that is sensitive to cue reliabilities

In a first step, a series of *t*-tests for significant differences from 0 was conducted on the single-cue weights of the control condition to examine whether all cues contributed to the depth judgments. Significant differences were obtained for shading, $t(9) = 6.81$, $p < .001$, texture, $t(9) = 9.22$, $p < .001$, and disparity weights, $t(9) = 7.51$, $p < .001$. Thus, all cues were taken into account when judging the cylinder's depth. In a second step, the single-cue weights of the control condition were compared using an ANOVA with the within-subject factor cue. A marginally significant effect was obtained, $F(2, 18) = 3.39$, $\varepsilon = .74$, $p < .10$, $f = 0.61$. As can be seen from Figure 6.2, the shading cue received the smallest weight and the disparity cue the largest. Afterwards, weights were calculated for the conditions with pairwise cue consistencies and a 2 × 3 ANOVA using condition and cue as within-subject factors was carried out to examine whether the weights that were assigned to the deviating cues in conditions with pairwise cue consistencies resemble the single-cue weights of the control condition. Such a result would have been predicted by simple additive cue combination models. By contrast, the analysis yielded a significant interaction of condition and cue, $F(2, 18) = 4.72$, $\varepsilon = .78$, $p < .05$, $f = 0.12$. Pairwise *t*-tests[6] between the single-cue

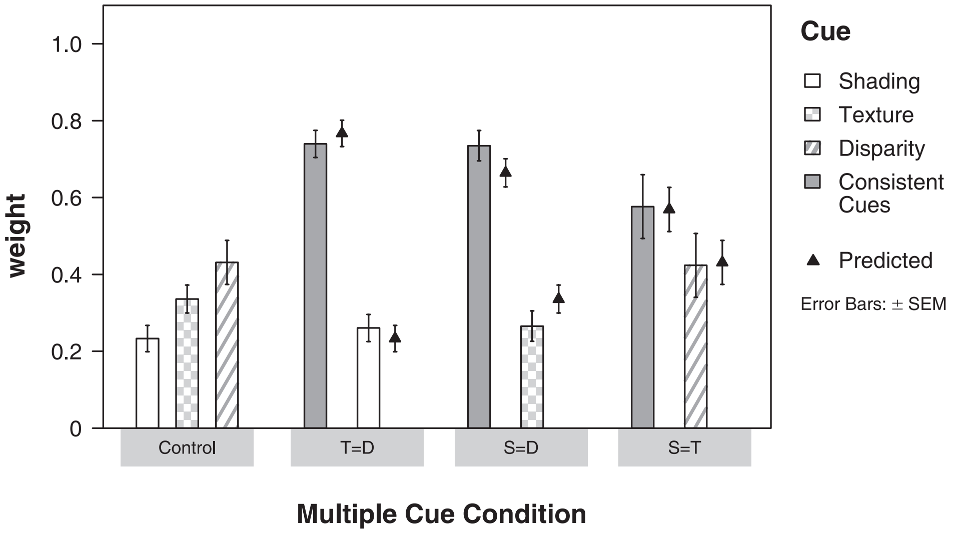


Figure 6.2. Perceptual weights for each cue or cue combination as a function of experimental condition. The characters S, T, and D denote the shading, texture, and disparity cue, respectively. The black triangles correspond to the predictions of the MLE model (see 6.3.2.2).

[6] Because multiple *t*-tests were performed, the Bonferroni correction was applied to diminish the probability of committing a type I error.

weights of each condition revealed a significant difference of the texture weight between both conditions, $t(9) = 5.18$, $p < .01$, but not for the shading, $t(9) = 1.20$, $p = .78$, or the disparity cue, $t(9) < 1$, respectively. Thus, the texture cue received a smaller weight as in the control condition when shading and disparity indicated a similar depth of the stimulus (see Figure 6.2; results for each observer are depicted in Appendix C).

In addition to single-cue weights, *d'* was calculated as an index of observer sensitivity for each experimental condition. An ANOVA using the within-subject factor condition (control condition and three conditions for each pairwise cue consistency) was conducted to examine whether *d'* was reduced in the conditions with pairwise cue consistencies as predicted by simple additive cue combination models. A significant main effect of condition was revealed, $F(3, 27) = 3.62$, $\varepsilon = .79$, $p < .05$, $f = 0.59$ (see Figure 6.3). Planned contrasts using the control condition as basis of comparison yielded significantly lower *d'* values for the conditions with consistent texture and disparity cues, $F(1, 9) = 6.04$, $p < .05$, as well as consistent shading and texture cues, $F(1, 9) = 8.37$, $p < .05$. Interestingly,

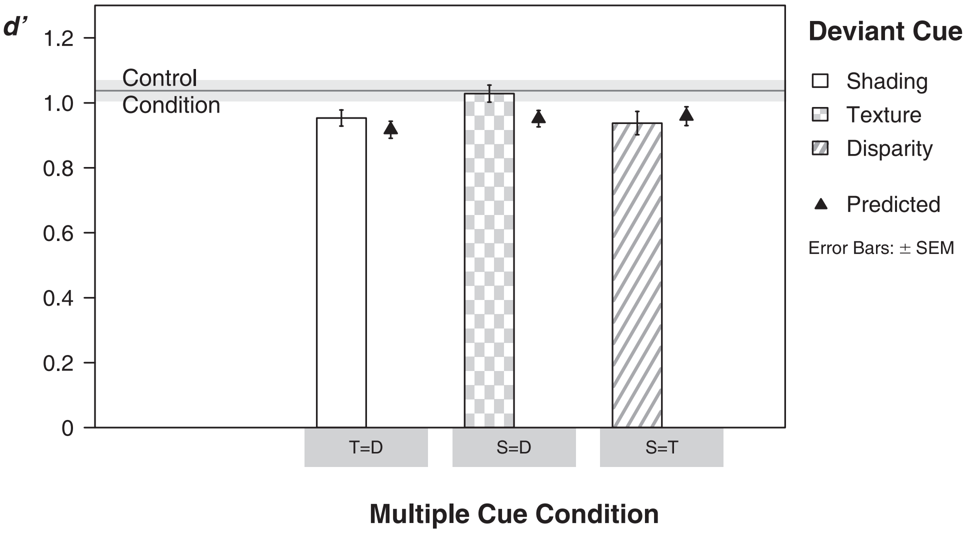


Figure 6.3. Observer sensitivity as indicated by *d'* for the conditions with pairwise cue consistencies. The result of the control condition is depicted as gray horizontal bar (mean *d'* ± standard error of the mean). The characters S, T, and D denote the shading, texture, and disparity cue, respectively. The black triangles correspond to the predictions of the MLE model (see 6.3.2.2).

the observers' sensitivity in the condition with consistent shading and disparity cues did not significantly differ from the control condition, $F(1, 9) < 1$. These results indicate that the observers improved their weighting of single cues when shading and disparity indicated a comparable depth of the stimulus. This reweighting allowed them to achieve a similar sensitivity as in the control condition (results for each observer are shown in Appendix C).

For the conditions with reduced reliabilities, a simple additive linear integration model that is sensitive single-cue reliabilities would predict a reduced weighting of the less reliable cue as compared to both other depth cues. This effect should statistically emerge in a significant condition by cue interaction on the empirical weights in separate 2 × 3 ANOVAs using the within-subject factors condition (control vs. reduced reliability condition) and cue. Indeed, a significant interaction was obtained for the reduced reliability of the shading cue, $F(2, 18) = 6.56$, $\varepsilon = 1.00$, $p < .01$, $f = 0.15$, the texture cue, $F(2, 18) = 7.66$, $\varepsilon = 1.00$, $p < .01$, $f = 0.15$, and the disparity cue, $F(2, 18) = 20.27$, $\varepsilon = .90$, $p < .001$, $f = 0.66$[7]. As

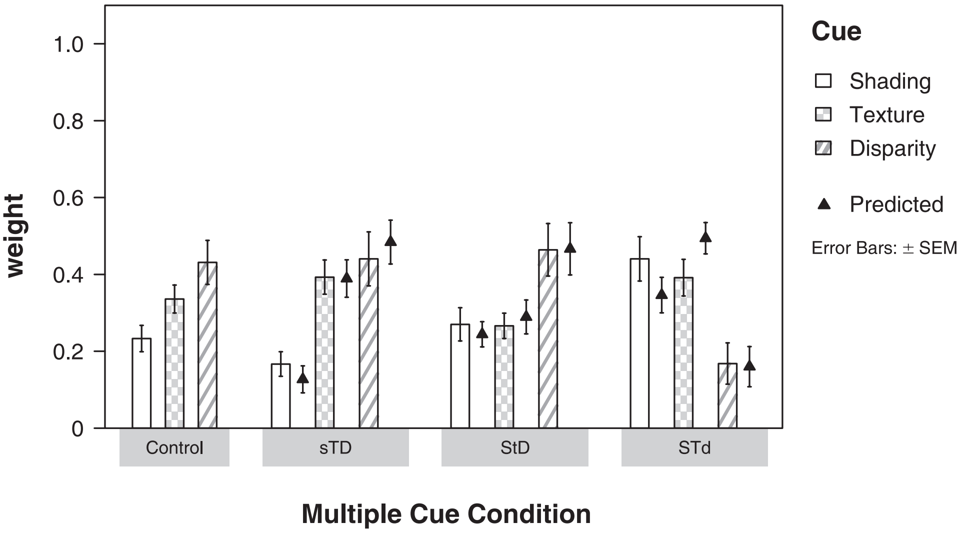


Figure 6.4. Perceptual weights for each cue as a function of experimental condition. The characters S, T, and D denote the shading, texture, and disparity cue, respectively, and reduced single-cue reliabilities are indicated by lower case letters. The black triangles correspond to the predictions of the MLE model (see 6.3.2.3).

[7] Moreover, marginally significant main effects of the cue factor were obtained in the ANOVAs on the empirical weights with reduced reliability of the shading, $F(2, 18) = 4.46$, $\varepsilon = .62$, $p < .10$, $f = 0.68$, and the texture cue, $F(2, 18) = 3.20$, $\varepsilon = .64$, $p < .10$, $f = 0.58$, respectively.

can be seen from Figure 6.4, reducing the reliability of a single cue led to a marked decrease of its weight in the multiple-cue observation task. Both other cues were upweighted correspondingly (results for each observer are depicted in Appendix C).

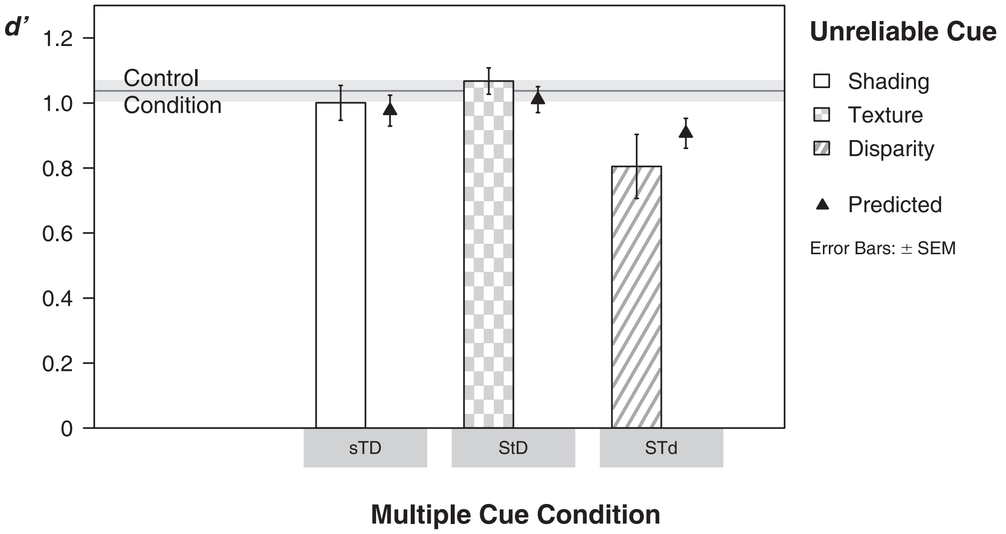


Figure 6.5. Observer sensitivity as indicated by *d'* for the conditions with reduced single-cue reliabilities. The result of the control condition is depicted as gray horizontal bar (mean *d'* ± standard error of the mean). The characters S, T, and D denote the shading, texture, and disparity cue, respectively, and reduced reliabilities are indicated by lower case letters. The black triangles correspond to the predictions of the MLE model (see 6.3.2.3).

Using a simple additive linear integration scheme, the reduced reliability of single cues should result in an increase of the percept's variability, thus reducing the observers' sensitivity *d'*. To test this assumption, an ANOVA using the within-subject factor condition (control condition and three conditions with reduced single-cue reliabilities) was conducted on the *d'* values. This ANOVA yielded a significant main effect of condition, $F(3, 27) = 5.45$, $\varepsilon = .51$, $p < .05$, $f = 0.70$. Planned contrasts using the control condition as basis of comparison only revealed significantly lower *d'* values in the condition with reduced reliability of the disparity cue, $F(1, 9) = 6.10$, $p < .05$. For the conditions with less reliable shading, $F(1, 9) = 1.51$, $p = .25$, or texture cues, $F(1, 9) < 1$, no significant difference from the control condition was obtained. Thus, only reducing the reliability of the disparity cue led to a marked decrease of *d'* (see Figure 6.5; results for each observer are shown in Appendix C). For both other cues, weights were adapted to take into account the reduced reliability but the cylinder's depth was judged as precisely as in the control condition.

6.3.2 Application of the MLE model

Overall, the empirical data roughly followed the qualitative predictions of an additive linear integration scheme that is sensitive to single-cue reliabilities. Perceptual weights proved to be stable in the conditions with pairwise cue consistencies and reduced *d'* values were observed with the exception of the condition with consistent shading and disparity cues. Reducing the reliability of single cues led to a downweighting of these cues and lowered the observer's sensitivity at least for comparably less reliable disparity information. Given these results, it would be interesting to quantitatively compare the empirical data to the predictions of the MLE model which is a special case of a linear additive integration scheme that takes into account single-cue reliabilities in an optimal fashion (Ernst & Bülthoff, 2004; Oruç et al., 2003).

6.3.2.1 Derivation of single-cue reliabilities

For the application of the MLE model to the current data, single-cue reliabilities are needed to predict perceptual weights as well as *d'* for the different experimental conditions. Unfortunately, these reliabilities cannot be directly calculated from the empirical data because unlike to previous research (e.g., Alais & Burr, 2004; Ernst & Banks, 2002; see also Experiment 1 and 2), no single-cue conditions were realized. However, these values can be estimated from the perceptual weights and *d'* of the control condition using certain assumptions that are directly related to the MLE framework.

First, the MLE model predicts an optimal integration of single cues into a combined estimate with respect to single-cue reliabilities. Thus, for the current

Experiment, the perceptual weights w_i can be assumed to be directly related to single-cue reliabilities r_i as follows:

$$w_\mathrm{S} = \frac{r_\mathrm{S}}{r_\mathrm{S} + r_\mathrm{T} + r_\mathrm{D}}; w_\mathrm{T} = \frac{r_\mathrm{T}}{r_\mathrm{S} + r_\mathrm{T} + r_\mathrm{D}}; w_\mathrm{D} = \frac{r_\mathrm{D}}{r_\mathrm{S} + r_\mathrm{T} + r_\mathrm{D}} \tag{6.1}$$

with

$$r_\mathrm{S} = \frac{1}{\sigma_\mathrm{S}^2}; r_\mathrm{T} = \frac{1}{\sigma_\mathrm{T}^2}; r_\mathrm{D} = \frac{1}{\sigma_\mathrm{D}^2} \tag{6.2}$$

The proportion of cue reliabilities could be determined using the estimated perceptual weights $\hat{w}_i$, but the precise values could only be calculated by taking *d’* into account. The observer’s sensitivity *d’* is defined as the ratio of the distance between the distributions of deep and shallow stimuli Δ_Depth divided by the pooled standard deviation σ of both distributions:

$$d' = \frac{\Delta_\mathrm{Depth}}{\sqrt{\sigma^2}} \tag{6.3}$$

The distance between the distributions of deep and shallow stimuli Δ_Depth was fixed to 20 cm in the current Experiment. The total variance σ^2 can be decomposed into the effective variance of the stimulus σ_stim^2 and the additional variance that is related to the imprecise estimation of single cues by the perceptual system $\sigma_\mathrm{percept}^2$ assuming both values to be uncorrelated.

$$\sigma^2 = \sigma_\mathrm{stim}^2 + \sigma_\mathrm{percept}^2 \tag{6.4}$$

Because the stimulus generation properties regarding the determination of single-cue depths are known (they were randomly sampled from a Gaussian distribution with σ = 20 cm), $\hat{\sigma}_\mathrm{stim}^2$ can be easily determined using the estimated perceptual weights $\hat{w}_\mathrm{i}$.

$$\hat{\sigma}_\mathrm{stim}^2 = \hat{w}_\mathrm{S}^2 \cdot 20^2\,\mathrm{cm}^2 + \hat{w}_\mathrm{T}^2 \cdot 20^2\,\mathrm{cm}^2 + \hat{w}_\mathrm{D}^2 \cdot 20^2\,\mathrm{cm}^2 \tag{6.5}$$

The empirical uncertainty of the perceptual system $\hat{\sigma}^2_{\text{percept}}$ can be determined by inserting the empirically estimated sensitivity $\hat{d}'$ and the stimulus variance $\hat{\sigma}^2_{\text{stim}}$ into equation 6.3. Solving it for $\hat{\sigma}^2_{\text{percept}}$ yields:

$$\hat{\sigma}^2_{\text{percept}} = \left(\frac{20\text{cm}}{\hat{d}'}\right)^2 - \hat{w}_{\text{S}}^2 \cdot 20^2\,\text{cm}^2 + \hat{w}_{\text{T}}^2 \cdot 20^2\,\text{cm}^2 + \hat{w}_{\text{D}}^2 \cdot 20^2\,\text{cm}^2 \quad (6.6)$$

Within the MLE framework, this variance is directly related to the (unknown) single-cue reliabilities r_{i} as follows:

$$\sigma^2_{\text{percept}} = w_{\text{S}}^2 \cdot \frac{1}{r_{\text{S}}} + w_{\text{T}}^2 \cdot \frac{1}{r_{\text{T}}} + w_{\text{D}}^2 \cdot \frac{1}{r_{\text{D}}} \quad (6.7)$$

Thus, for the current experiment, single-cue reliabilities could be determined from the empirical weights $\hat{w}_{\text{i}}$ and the estimated observer sensitivity $\hat{d}'$ of the control condition using equations 6.6 and 6.7.

6.3.2.2 Fit of the empirical data to the predictions of the MLE model in conditions with pairwise cue consistencies

In its standard form, the MLE framework does not take cue interactions into account. Thus, consistent cues should be weighted according to the sum of both single-cue weights, which, in turn, should be related to their reliabilities according to equation 6.1. A statistical comparison of the empirical weights to the weights that were predicted according to this logic is presented in Table 6.2. The results resemble the analyses in section 6.3.1 by demonstrating that the MLE predictions fit to the empirical weights except for the condition with consistent shading and disparity cues as indicated by a significant data × cue interaction in this condition. In this case, the empirical weight of the deviant texture cue was significantly reduced and fell short of the predicted value (see black triangles in Figure 6.2 illustrating the predicted weights).

Table 6.2. Differences between predicted and empirical weights and *d'* values for conditions with pairwise cue consistencies. With respect to the perceptual weights, 2 × 2 analyses of variance (ANOVA) using the within-subject factors data (A: empirical vs. predicted) and cue (B: consistent vs. deviant) were conducted and the main effect of cue B as well as the interaction term A × B are reported for each condition. To compare predicted and empirical *d'* values, pairwise *t*-tests were conducted.

		Weights			*d'*	
Condition	Effect	*F*(1, 9)	*p*	*f*	*t*(9)	*p*
Texture = Disparity	A				1.36	.21
	B	59.28	< .001	2.40		
	A × B	1.43	.26	0.05		
Shading = Disparity	A				3.46	< .01
	B	28.50	< .001	1.67		
	A × B	26.81	< .001	0.15		
Shading = Texture	A				0.53	.61
	B	1.09	.32	0.34		
	A × B	0.07	.80	0.02		

Note. With respect to the ANOVAs, the main effect A could not be specified as the weights summed up to 1 for both conditions. Cohen's *f* is reported as an effect size estimate.

With respect to the observer sensitivity *d'*, some additional considerations are necessary to derive predictions of the MLE model. Because no interactions between cues are included in the simple MLE framework, it can be assumed that the uncertainty $\hat{\sigma}^2_{\text{percept}}$ of the perceptual system remains constant. That is, it can be estimated from the control condition. The effective variance of the stimulus $\hat{\sigma}^2_{\text{stim}}$, however, is reduced when two cues indicate a similar depth of the stimulus. This is due to the reduction of independently sampled estimates of the stimulus depth and it can be calculated using the following equations depending on which pair of cues indicated a consistent depth.

$$\begin{aligned}
\hat{\sigma}^2_{\text{stim}} &= (\hat{w}_{\text{T}} + \hat{w}_{\text{D}})^2 \cdot 20^2\,\text{cm}^2 + \hat{w}^2_{\text{S}} \cdot 20^2\,\text{cm}^2; \\
\hat{\sigma}^2_{\text{stim}} &= (\hat{w}_{\text{S}} + \hat{w}_{\text{D}})^2 \cdot 20^2\,\text{cm}^2 + \hat{w}^2_{\text{T}} \cdot 20^2\,\text{cm}^2; \\
\hat{\sigma}^2_{\text{stim}} &= (\hat{w}_{\text{S}} + \hat{w}_{\text{T}})^2 \cdot 20^2\,\text{cm}^2 + \hat{w}^2_{\text{D}} \cdot 20^2\,\text{cm}^2
\end{aligned} \tag{6.8}$$

Using equations 6.5 and 6.8, $\hat{\sigma}^2_{\text{stim}}$ and $\hat{\sigma}^2_{\text{percept}}$ were calculated for each condition with pairwise cue consistencies. Afterwards, these values were used to derive the predicted *d'* value by inserting $\hat{\sigma}^2_{\text{stim}} + \hat{\sigma}^2_{\text{percept}}$ into equation 6.3. These values are depicted in Figure 6.3 by black triangles and predicted and empirical *d'* values were compared using pairwise *t*-Tests with the results displayed in Table 6.2. The predictions of the MLE model seemed to hold for consistent texture and disparity as well as shading and texture cues, respectively, but when shading and disparity indicated a comparable depth of the stimulus, the empirically estimated observer sensitivity was significantly larger than predicted by the MLE model.

6.3.2.3 Fit of the empirical data to the predictions of the MLE model in conditions with reduced single-cue reliabilities

For conditions with reduced single-cue reliabilities, the predictions of the MLE model cannot be directly calculated from the weights and *d'* of the control condition because single-cue reliabilities can be expected to change substantially. Thus, they have to be estimated from the empirical data for each condition with reduced reliabilities separately. By assuming that only the reliability of the cue with added perceptual noise changes from the respective value in the control condition, single-cue reliabilities can be estimated using equations 6.6 and 6.7. The reliabilities of the noisefree cues were set to the values that were derived from the control condition (see 6.3.2.1) and the reliability of the remaining cue was fitted to reproduce the *d'* that was empirically observed in the respective condition by using a least squares fit. Afterwards, predicted weights and *d'* were calculated separately for each condition (see black triangles in Figures 6.4 and 6.5).

Empirically derived values were compared to the predictions of the MLE model with the results displayed in Table 6.3. Significant differences were only obtained for the observer sensitivity in the condition with reduced reliability of the texture cue. In this case, the empirical *d'* was significantly larger than predicted by the MLE framework.

Table 6.3. Differences between predicted and empirical weights and *d'* values for conditions with reduced single-cue reliabilities. With respect to the perceptual weights, 2 × 3 analyses of variance (ANOVA) using the within-subject factors data (A: empirical vs. predicted) and cue (B: shading, texture, disparity) were conducted and main effects as well as the interaction term are reported for each condition. To compare predicted and empirical *d'* values, pairwise *t*-tests were conducted.

		Weights				*d'*	
Less reliable cue	Effect	*F*(2, 18)	ε	*p*	*f*	*t*(9)	*p*
Shading	A					1.28	.23
	B	8.47	.61	< .05	0.90		
	A × B	0.95	.94	.40	0.09		
Texture	A					2.41	< .05
	B	3.64	.61	.08	0.62		
	A × B	0.71	.89	.49	0.06		
Disparity	A					1.54	.16
	B	10.60	1.00	< .001	0.78		
	A × B	1.46	1.00	.26	0.21		

Note. With respect to the ANOVAs, the main effect A could not be specified as the weights summed up to 1 for both conditions. Huynh-Feldt ε values and Cohen's *f* are reported for each effect.

6.3.3 Subjective difficulty and response times

After the Experiment, observers were asked to rate the difficulty of estimating the depth of one prototypical multiple-cue stimulus of each condition on a 4-point Likert scale ranging from high (1) to low (4). Subjective difficulties are displayed in Table 6.4 along with response times as a function of experimental condition[8]. Overall, responses were given very fast and after the Experiment, most observers reported that they quickly came to a decision for the particular stimulus and anticipated the stimulus offset to press the corresponding key. The subjective difficulty seemed to be higher for conditions with reduced single-cue reliabilities.

[8] The response time was defined as the interval from the stimulus offset to the observer's key press. Response times larger than 3 s were excluded. Such slow responses occurred in only 0.14% of all trials on average.

Moreover, response times were slightly larger in these conditions, too. When collapsing data across the three conditions with pairwise cue consistencies as well as reduced reliabilities, respectively, it turned out that the subjective difficulty, $t(9) = 2.95$, $p < .05$, as well as the response times increased, $t(9) = 2.75$, $p < .05$, when single cues had a lower reliability as compared to the control condition. By contrast, no such differences were observed for the conditions with pairwise cue consistencies: $t(9) < 1$, for subjective difficulties and response times, respectively.

Table 6.4. Subjective difficulty and response times for the different experimental conditions. Additionally, values were separately averaged for all conditions with pairwise cue consistencies and reduced single-cue reliabilities, respectively.

		Subjective difficulty (1=high, 4=low)		Response time (ms)	
Condition		*M*	(*SD*)	*M*	(*SD*)
Control	STD	3.10	(0.88)	172.24	(62.28)
Pairwise consistencies	S=T	2.80	(0.79)	172.04	(62.16)
	S=D	2.90	(0.99)	174.06	(64.63)
	T=D	3.40	(0.70)	170.17	(56.09)
	Total	3.03	(0.58)	172.09	(60.25)
Reduced Reliability	sTD	2.40	(0.84)	182.16	(67.69)
	StD	2.20	(0.92)	179.04	(67.23)
	STd	2.20	(1.14)	179.58	(66.34)
	Total	2.27	(0.34)	180.26	(65.59)

Note. The characters S, T, and D denote the shading, texture, and disparity cue, respectively, and reduced single-cue reliabilities are indicated by lower case letters.

After the Experiment, observers were additionally asked to indicate whether they adopted their depth estimation strategy to certain aspects of the prototypical stimuli that were presented for each condition. Interestingly, all reported that they changed their strategy when the disparity information was less reliable than in the control condition by trying to ignore this “noisy” cue. Less than half of the observers followed a similar strategy when the reliability of the shading or texture cue was reduced. Regarding the pairwise cue consistencies, only one observer in each condition reported a conscious change of her depth estimation strategy. However, the corresponding descriptions were relatively vague so they might have been more related to specific attributes of the prototypic stimulus instead of revealing a reliable strategy for the respective condition.

6.4 Discussion

The current experiment allowed for a qualitative test of the predictions of an additive linear integration scheme that is sensitive to single-cue reliabilities. Furthermore, under certain assumptions, the empirical data could also be compared to the quantitative predictions of the MLE model. In contrast to previous work that mostly examined the integration of two separate visual cues (e.g. Hillis et al., 2004; Jacobs, 1999; Knill & Saunders, 2003), three cues were manipulated in the current study to test the impact of pairwise cue interactions. Furthermore, no single-cue conditions were realized in this experiment because there might be qualitative differences between the perceptual processing in single-cue and multiple-cue conditions. To allow for an estimation of single-cue weights, a multiple observation task was realized (see section 5.1; Berg, 1989, 1990; Lutfi, 1995).

6.4.1 Adequacy of the MLE model

Overall, observers put the largest weight on the disparity and the smallest on the shading cue in the control condition. This pattern of results is roughly comparable to Experiment 3, where the disparity cue did also have the largest impact on the observers' depth judgments. In the current experiment, however, the shading and texture cue were slightly upweighted in comparison to Experiment 3. All observers reported having noticed variations in the three depth cues and the average weights indicate that all cues actually contributed to the observers' depth estimates.

The empirical data (weights and *d'*) of the conditions with pairwise cue consistencies as well as reduced single-cue reliabilities closely followed the predictions of the MLE model for most conditions. Comparable results using a quantitative test of the MLE model have been reported for the integration of stereo and texture cues in slant discrimination (Hillis et al., 2004; Knill & Saunders, 2003; Saunders & Backus, 2006), and depth estimation (Adams & Mamassian, 2004). However, some deviations from the MLE predictions were also observed in the current study. Most importantly, the texture cue was downweighted when shading and disparity indicated a similar depth of the stimulus. This led to an increase of the observer's sensitivity in this condition above the predicted value of the MLE model. Reducing the reliability of the texture cue in yet another experimental condition induced a downweighting of the respective cue as predicted by the MLE model, which, however, did not result in a decrease of *d'*. These data indicate that a deviation of the texture cue with respect to an increased discrepancy from the other cues or decreased reliability led to non-linearities in the process of cue integration. Thus, cue interactions seem to be very specific and do not necessarily exist for all cues. With respect to the current study, for example, it might be suggested that shading and texture as well as disparity and texture are processed independently from each other in separate modules. Shading and disparity cues, however, seem to interact in the process of depth estimation.

An independent processing of shading and texture cues has already been suggested on the basis of several studies at near threshold level examining the processing of luminance-modulated and contrast-modulated gratings as models for shading and texture cues. Schofield and Georgeson (1999), for example, did not find facilitation or sub-threshold summation between barely detectable luminance-modulated and contrast-modulated gratings. Furthermore, Georgeson and Schofield (2002) observed a transfer of aftereffects between channels, but the identity of both cue patterns was not lost during this integration. These results further substantiate an independent processing of shading and texture cues. Regarding texture and disparity, a strong interaction of both cues was originally suggested as texture might help calibrating the stereo cue regarding viewing direction and distance between observer and fixation point. This calibration is required to allow for absolute instead of relative depth judgments. Frisby et al. (1995) conducted a series of experiments to reveal whether stereo information is indeed calibrated using texture cues and completely failed to find evidence for this hypothesis. A linear additive integration of stereo and texture cues without cue interactions was also reported by Hillis et al. (2004) and Knill and Saunders (2003) for judgments of surface slant. Thus, texture and disparity information seem to be processed independently which fits to the results of the present experiment.

Such pattern of highly specific cue interactions was also reported for depth perception at near threshold levels. For example, Ichikawa and colleagues (Ichikawa, Saida, Osa, & Munechika, 2003) used gratings whose depth was defined by disparity, motion parallax, or monocular shape cues. The cues could be consistent (in-phase) or discrepant (out-of-phase or different spatial frequencies). It turned out that the sensitivity enhancement when combining single cues was most pronounced and even larger than predicted by probability summation[9] when consistent disparity and motion parallax cues were present in the scene (q.v., Bradshaw & Rogers, 1996). For all other cue combination, however, the observed *d'* values were roughly comparable to the predictions of the MLE model or fell even short these values. The authors therefore concluded that the visual system seems to integrate depth information from different cues in different ways which also applies for the current experiment.

[9] The predicted *d'* of two combined cues when assuming probability summation of independent cue specific channels is equal to the square root of the quadratic sum of both single-cue *d'* values (McMillan & Creelman, 2005, p. 158). This corresponds to an additive integration theme disregarding cue interactions. Thus, it shares important features with the MLE framework (see Gepshtein et al., 2005, p. 1014 f.).

6.4.2 Interactions of shading and disparity

As outlined above, a simple additive cue integration scheme could not account for the empirical data of the present study when the texture cue deviated from shading and disparity with respect to consistency or reliability. Such pattern of results might be due to an interactive processing of shading and disparity information. The integration of stereo and shading cues was already examined in one of the first studies on visual cue integration (Bülthoff & Mallot, 1988). In this study, observers judged the depth of an ellipsoid that was defined by stereo disparity (provided by localized edges), stereo shading (depicted by a shifted luminance gradient in each monocular image), monocular shading, or a combination of these cues. Overall, perceived depth increased with the availability of cues and was almost veridical when all cues were present in the scene. In case of conflicting cues, stereoscopic disparity completely determined the percept (cue vetoing). Thus, this early study already suggests a nonlinear integration of stereo and shading cues.

Much work on the perception of complex and naturalistic objects was conducted by Koenderink and colleagues using surface attitude settings (cf., Koenderink et al., 1992). In this method, which is based on earlier work by Mingolla and Todd (1986), observers have to adjust a small gauge figure to reproduce the local surface slant and tilt at numerous positions in the image plane. These adjustments can subsequently be used to reconstruct the perceived object. In one study, Doorschot, Kappers, and Koenderink (2001) used this method to examine the influence of disparity and shading on visual shape perception. Photographs of two plastic torsos were presented monocularly or stereoscopically with an eye base of 7 or 14 cm. Additionally, a shading pattern was generated by illuminating the torso from one of three directions. Overall, shading and stereo influenced the perceived structure of the objects in an additive fashion, thus conflicting with the nonlinear integration that was observed in the present experiment. It has to be noted, however, that shape consistency was clearly the most relevant factor in the study by Doorschot et al. (2001) accounting for about

94% of the variance. Thus, shading and stereo only had a small additional effect on shape perception which complicates the detection of a potential cue interaction.

A recent study by Vuong et al. (2006) is highly relevant for interpreting the results of Experiment 4. These authors were interested in the potential cooperation of disparity and shading cues for surface interpolation. Therefore they measured the reliability of surface judgements for each cue separately and for their combination, thus closely following the experimental logic of most studies on the MLE integration scheme (see Ernst & Bülthoff, 2004, for a review). In both experiments that were conducted by Vuong and colleagues, the precision of depth estimates based on the disparity cue was larger than the reliability of the shading cue. Furthermore, adding shading to the display led to an increase of depth judgements' precision as compared to displays depicting the disparity cue alone. Most importantly, this improvement of multiple-cue reliability was larger than predicted by the MLE model. Thus, this result is fully comparable to the present study where *d'* was significantly larger than predicted by the MLE framework when shading and disparity indicated a similar depth of the cylinder. A comparable result was also observed for the condition with a less reliable texture cue. Potentially, this reduction of reliability directed the observer's attention more strongly to shading and disparity cues (as also verified by changes in single-cue weights). In turn, these two cues could be processed cooperatively, thus enhancing the observer's sensitivity above the prediction of the MLE model.

6.4.3 Conclusions

Experiment 4 demonstrated that specific interactions between visual depth cues influence the process of integration. Although the MLE model accounted for a large amount of data in the current experiment, it failed to predict the observers' response pattern for conditions with deviating texture information. The Bayesian model that was outlined in section 2.5 and successfully utilized in Experiments 1 and 2 could not be directly tested in the current experiment because single-cue reliabilities were not independently measured. It is interesting to note, however, that the Bayesian model might account for the current data in several ways: 1) Within this model, multiple-cue reliabilities can be higher than predicted by a simple additive integration rule – as was found for this experiment. 2) Because multiple cues are not assumed to be perfectly fused, top-down effects regarding the differential direction of attention can be easily included. Such a reorientation of attention might account for the unexpectedly high *d'* in the condition with a reduced reliability of the texture cue in the current experiment. 3) The coupling prior in the Bayesian model allows for modelling highly specific cue interactions. Thus, two modules can be assumed to independently process depth information or they can be highly coupled. With respect to the current study, it might be supposed that shading and disparity are processed cooperatively that is, they are coupled to a stronger degree than all other cue pairings. Overall, the Bayesian model including a coupling prior might also account for results in conditions where the MLE model failed to predict the empirical response pattern.

7. General discussion

7.1 Factors influencing sensory integration

The human perceptual system seems to effortlessly integrate different senses and numerous cues within one sense into a highly monolithic percept. These mainly unconscious processes continuously provide robust information about the surrounding world. From an ecological point of view, it seems likely that sensory integration is sensitive to single-cue characteristics and multisensory constraints in the real world. For example, to optimally use the information of different cues, the perceptual system should integrate them with respect to their reliability. Moreover, multisensory signals from a common external event or object will often be spatially and temporally aligned and provide consistent estimates of the respective physical property that has to be judged. It seems reasonable to assume that such information affects sensory integration to allow for a robust fusion of multiple cues. These aspects will be discussed in the following sections by relating the empirical data of the current study to the available body of research from other sources.

7.1.1 Signal reliability

The reliability of individual signals strongly affects their integration into a combined estimate. Signals that can be estimated with high precision were repeatedly shown to be weighted to a larger degree than relatively unreliable signals (see review of empirical studies in sections 2.4.1 and 2.4.2). This reliability-sensitive cue weighting has been demonstrated for the combination of visual and haptic cues (Bresciani et al., 2006; Ernst & Banks, 2002; Helbig & Ernst, 2007), audiovisual signals (Alais & Burr, 2004; Andersen et al., 2004), proprioceptive and visual information (van Beers, Sittig, & Denier van der Gon, 1999), and auditory and tactile cues (Bresciani & Ernst, 2007). Furthermore, such processing of differentially reliable cues was shown in the intramodal case for the integration of visual cues in depth estimation (Jacobs 1999; Young et al., 1993), slant perception (Hillis et al., 2004; Knill & Saunders, 2003; Saunders & Backus, 2006), and for the integration of haptic cues in shape estimation (Drewing & Ernst, 2006). Taken together, individual cues are weighted according to their reliability in the intersensory as well as the intramodal case. Thus, this integration rule seems to reflect a general principle of how the central nervous system combines differentially reliable signals.

Single-cue reliability was also shown to affect the combination of shading, texture, and motion cues in Experiment 1 and 2. In these cases, however, pairwise cue consistencies were also taken into account. The empirical data of both experiments could well be described by an extended Bayesian modal that relied on single-cue reliabilities and multiple-cue consistencies (see section 2.5). A more direct demonstration of the influence of single-cue reliability on the integration of visual depth cues was provided by Experiment 4. In this case, the hemicylinder's depth that had to be estimated by the observers was defined by shading, texture, and stereoscopic disparity. Corrupting one of these cues by added perceptual noise led to a marked downweighting of this cue in the combined percept. The degree of downweighting was roughly proportional to the weight of the respective cue in the control condition. Thus, cues that strongly influenced the combined estimate were substantially downweighted when their reliability was reduced. This result extends findings of former studies on intrasensory integration that did not manipulate single-cue reliabilities (e.g., Jacobs 1999) or used potentially correlated depth cues (e.g., Young et al., 1993).

7.1.2 Spatial and temporal proximity

In addition to signal reliability as a property of individual cues, spatial and temporal proximity of several signals are necessary for initiating their integration. Much research on the latter aspects has been conducted by using audiovisual localization tasks. When observers have to indicate the location of an auditory event that is accompanied by a synchronously delivered visual signal from a slightly different spatial location, the response is typically shifted into the direction of the visual event (Bermant & Welch, 1976; Bertelson & Radeau, 1981; Thurlow & Rosenthal, 1976). This effect has been named “ventriloquism” because the spectacular illusion of a speaking puppet that can be generated by experienced ventriloquists also relies on a visual bias of auditory location. Much research has focused on factors influencing the amount of capture that is generated by presenting spatially discrepant visual events that have to be ignored when judging the location of the auditory signal. Increasing the spatial separation between both signals reduced the amount of ventriloquism (Bertelson & Radeau, 1981; Jack & Thurlow, 1973; Thurlow & Jack, 1973). Thus, the impact of visual signals on the perceived location of auditory events is inversely related to the spatial conflict between both channels. Comparable results were also obtained for varying temporal discrepancies between visual and auditory signals. When both signals were delivered in a completely asynchronous manner, no ventriloquism occurred (Bertelson & Aschersleben, 1998). When introducing slight, but systematic temporal differences, the visual signal still affected the localization of auditory events. However, the amount of this effect dropped sharply as the temporal discrepancy grew (Jack & Thurlow, 1973; q.v., Slutsky & Recanzone, 2001).

The influence of spatial proximity on the process of intersensory integration is not limited to audiovisual localization tasks. Gepshtein and colleagues (2005) asked observers to judge the distance of two virtual planes that were presented either visually, haptically, or in both modalities. For bimodally presented planes, visual and haptic signals were spatially coincident or separated by 30, 60, or 90 mm. It turned out that discrimination performance was maximal when both signals were presented at the same location. In this case, the reliability of the combined estimate could be successfully predicted from the unimodal conditions by using the simple MLE model. However, discrimination accuracy dropped continuously with increasing spatial offset and largely resembled the values of the unimodal tasks when both signals were separated by 90 mm. Thus, an improvement of perceptual precision within sensory integration depends on spatial proximity.

An automatic integration of visual and auditory signals with respect to their temporal occurrence was recently demonstrated by Fendrich and Corballis (2001). Observers had to estimate the time point when a visual flash or an auditory click occurred. In addition to the relevant event, a signal was presented in the respective irrelevant channel that could be temporally displaced. It turned out that the estimated temporal position was significantly influenced by the irrelevant signal. Thus, events that occurred in temporal proximity were likely to be integrated into a combined percept but this integration seems to be limited to small amounts of temporal discrepancy (cf. Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Bresciani et al. (2005) showed that the number of auditory beeps affected the perceived number of tactile taps only when both signals were presented in close temporal proximity and comparable results were also reported for audiovisual stimuli by Shams et al. (2002). However, the temporal window within which sensory signals are integrated seems to be flexible within certain limits. In a recent study by Navarra et al. (2005), participants had to judge the temporal order of brief visual and auditory signals (cf., Hirsh & Sherrick, 1961) while viewing speech or music video clips. When the auditory channel of the video was delayed by 300 ms, participants required larger discrepancies in order to judge the relative temporal position of visual and auditory signals correctly. This did not occur when the delay was increased to 1000 ms. Thus, the temporal window of audiovisual integration seems to be flexible when the observer is continuously confronted with slightly asynchronous signals.

Taken together, signals from different modalities are likely to be integrated when they are spatially and temporally coincident. However, this integration does not follow an all-or-none law. Instead, the amount of integration gradually declines as signals become spatially or temporally discrepant (cf., Wallace et al., 2004). Interestingly, this form of integration was already described by the proximity rule of Gestalt-Psychology several decades ago (Wertheimer, 1923).

7.1.3 Cue consistency

Under certain conditions, it seems unlikely that sensory signals are uniformly integrated into a combined estimate even when they are spatially and temporally coincident. Sensory integration should not occur or the amount of fusion should gradually decrease when these signals are inconsistent or supposed to be unrelated. A first empirical result pointing into this direction was reported by Gepshtein and Banks (2003). In this study, observers had to judge the distance of two virtual planes by using touch and vision. Whereas most of their data could well be described by referring to the MLE framework, the observed reliability in the visual-haptic condition was substantially smaller then predicted by the MLE-model. This deviation seemed to be related to the experimentally generated conflict between visual and haptic cues and was substantially reduced when the analysis was confined to conditions with small discrepancy between both signals. Thus, cue consistency seems to be an additional factor that determines the degree of multisensory integration. This aspect might be even more critical when combining several cues within one modality because in this case, inconsistencies can occur easily and are not restricted to a somewhat artificial laboratory setting. Consider, for example, one has to estimate the distance of several vehicles ahead when driving on a straight road. Several depth cues could be used to serve this purpose: Shading and texture gradients, linear perspective, or motion induced depth cues. Now suppose that the marking lines' distance changes suddenly. In this case, the output of a depth-from-texture module might become inconsistent to the other depth cues. In such situation, it would be important to downweight this cue in comparison to all other cues irrespective of the individual cue's reliability. Subsequently, the interpretation of this deviating cue should be adjusted to match the estimates from other depth cues (cf., Atkins et al., 2003).

Experiment 1 and 2 provided evidence for such a consistency based cue weighting in a visual depth estimation task. A simple MLE model that did not account for pairwise consistencies of shading, texture, and motion cues, largely failed to predict the empirical data. By contrast, an extended Bayesian model incorporating a coupling prior provided a better fit to the empirical estimates for most observers even when taking the larger complexity of this model into account. The results of Experiment 1 also indicate that prior knowledge about cue consistencies might influence this flexible reweighting. Especially the weight of the shading cue was considerably reduced in this experiment when both other cues indicated a comparable depth of the cylinder. This may be related to former perceptual experiences where shading differed from other cues because of its dependency on lighting conditions.

The generalizability of these results was tested in Experiment 4 using a slightly different setup. In this case, shading, texture, and stereoscopic disparity were used as depth cues and it turned out that most of these results could be explained by a simple additive integration rule. However, when shading and stereo indicated a similar depth of the virtual object, the dissenting texture cue was downweighted but the observers' sensitivity remained stable. Such a result cannot be explained by assuming an independent processing of single cues and might indicate that shading and stereo are processed cooperatively. This interpretation fits to former studies showing a non-linear integration of stereo and shading cues in depth estimation (e.g., Bülthoff & Mallot, 1988; Vuong et al., 2006) and indicates that cue interactions can be highly specific. Obviously, the visual system integrates depth information from different cues in different ways (cf., Ichikawa et al., 2003). At least with respect to some cues, consistency seems to play an important role. The law of similarity of Gestalt-Psychology (Wertheimer, 1923) might be interpreted as an early conception of this principle.

7.2 Robust integration

All above mentioned parameters that affect the integration of multiple cues across or within sensory systems can be supposed to contribute to robust fusion. This concept describes the ability of sensory systems to provide relatively stable estimates of physical properties in complex and variable situations (cf., Landy et al., 1995; Maloney & Landy, 1989). For example, the judged distance of objects does not dramatically change when one of the eyes is closed. Thus, although depth cues (disparity and vergence) completely drop out in this situation, a stable perception of the current scene is maintained. This might be realized by using cue consistency to identify dissenting or missing depth estimates that should be vetoed or downweighted.

The simple MLE model is only robust for situations with one highly reliable cue that allows for a veridical perception of the current scene. However, when several cues with differing reliabilities are present in the scene, this model disregards cue consistencies and assumes an integration which is solely based on cue reliability. This situation is depicted in Figure 7.1 where two cues are relatively consistent (Cue 1 and 3) and one highly reliable cue differs from these estimates. According to the predictions on the MLE model, the combined estimate linearly follows the estimate of Cue 2 and has a constant variance across all possible values of this dissenting cue (see section 2.4). The predictions of the Bayesian model that takes cue interactions into account (CMLE model, see section 2.5) also fall on a straight line which, however, has a much smaller slope. Thus, the combined estimate is less affected by the discrepant information of Cue 2. Moreover, the variance of this combined estimate increases as a function of the dissenting cue's discrepancy from the other signals (see gray areas in Figure 7.1, panel A).

Although the extended Bayesian model might be regarded as providing more robust estimates of the given physical property as the MLE model, its combined measure still linearly depends on the value of Cue 2. A different prediction is displayed as dashed curve in panel B of Figure 7.1. In this case, the influence of Cue 2 on the combined percept is gradually decreased as a function of this cue's discrepancy from the other two cues. Such relationship can be mathematically described by the influence curve (Hampel, 1974) and might play an important role in the robust integration of multiple cues. A similar function was obtained for the integration of discrepant visual and auditory cues (Roach et al., 2006) and for varying amounts of spatial and temporal discrepancies in intersensory perception (e.g., Gepshtein et al., 2005; Shams et al., 2002).

Interestingly, the variance of the combined estimate within the extended Bayesian model provides information about the amount of discrepancy between the cues. Given that observers can assess this variance, it might be supposed that top-down processes come into play when this variance becomes too large. In this case, attentional processes might lead to a gradual downweighting of the dissenting cue to stabilize the combined percept. Such processing would also be more plausible than the robust estimation model that was proposed by Landy et al. (1995, p. 394 f.). They suggested that discrepant cues might be identified and downweighted on the basis of a statistical outlier rejection procedure such as the trimmed mean technique. These approaches typically eliminate a certain proportion of extreme values from the sample (e.g. 5%) before calculating the relevant statistics (Maronna, Martin, & Yohai, 2006, p. 31 f.). Consequently, a large number of data points is needed to utilize this technique. With respect to cue integration in visual depth perception, only a limited number of cues is available (e.g., two to six, Knill, 2007, p. 14 f.), thus impeding a direct transfer of robust statistical estimation procedures to the problem of sensory integration.

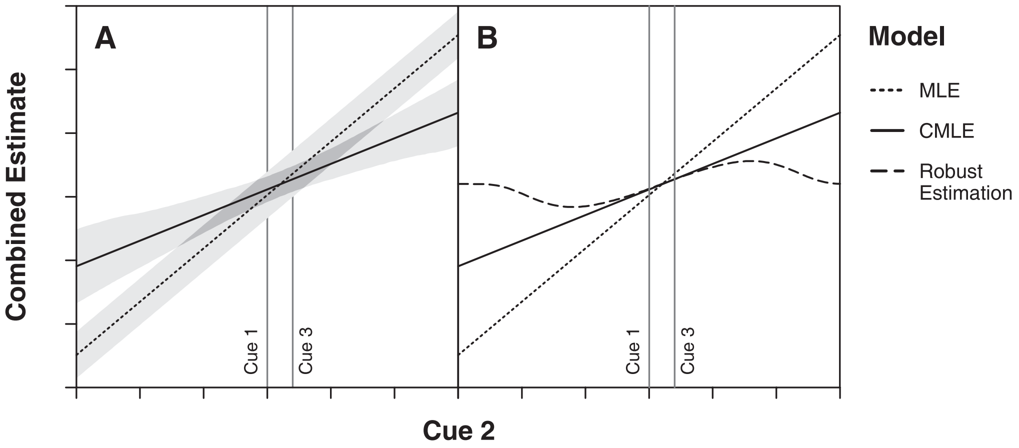


Figure 7.1. Predicted multiple-cue estimates of the simple MLE model and the extended Bayesian model that takes cue consistencies into account (CMLE model) for the integration of two consistent, but relatively unreliable cues (Cue 1 and 3) as a function of one highly reliable Cue 2. In panel A, the standard deviation of the combined estimate is depicted by the gray areas. In panel B, a third cue integration rule is additionally shown where the influence of Cue 2 gradually decreases as a function of its discrepancy from the other two cues.

7.3 Neural computation

Behavioral studies on multimodal and intrasensory integration suggest that single-cue reliability, spatial and temporal proximity, and multi-cue consistency are taken into account when estimating specific stimulus properties. Bayesian inference seems to be a powerful approach to model these influences on perception and action but it is yet unclear how such models might be implemented on the neural level. Nevertheless, several suggestions have been made in recent years on how sensory information might be represented and processed by neural networks (see Pouget, Dayan, & Zemel, 2003, for a review).

Central to all Bayesian approaches is the probabilistic representation of information. This can be achieved most easily by encoding probability information in the spiking rate of single neurons. The spiking rate would thus be proportional to the probability of specific input values. However, such an approach has several disadvantages. On the one hand, the reliability of such a coding scheme is highly limited due to neural variability. On the other hand, each neuron selectively processes only one specific stimulus attribute, thus complicating the integration of several different stimulus properties.

A different scheme that relies on population codes instead of the activity of single neurons seems to be more appropriate for encoding and processing probabilistic information in the neural system (Dayan & Abbott, 2001, p. 97 ff.). Neurons constituting such population typically have different but overlapping selectivities and the activation pattern $\mathbf{r}$ of such a population of n neurons can be used to effectively encode a given stimulus property s. This encoding is illustrated on the left side of Figure 7.2. Panel A shows the idealized bell-shaped tuning curves of six exemplarily chosen neurons. For example, these cells might represent neurons in the primary visual cortex responding to different orientations of a simple stimulus (Hubel & Wiesel, 1962). When a large number of such cells encode a specific stimulus, a response pattern similar to panel B might result. These neurons were ranked according to their preferred stimulus and their spiking activity was corrupted by Poisson noise which seems to be a good approximation of neural noise in the visual cortex (Tolhurst, Movshon, & Dean, 1983). This neural response pattern represents the likelihood of a neural population activity **r** given a stimulus *s*. The brain now has to estimate the posterior probability distribution $p(s \mid \mathbf{r})$ from the neural response pattern to identify the given stimulus. This might be achieved by applying the Bayes theorem

$$p(s \mid \mathbf{r}) = \frac{p(\mathbf{r} \mid s) \cdot p(s)}{p(\mathbf{r})} \tag{7.1}$$

with $p(\mathbf{r} \mid s)$ representing the likelihood function and $p(s)$ the prior. For independent Poisson neural variability (see Figure 7.2, panel B), $p(s \mid \mathbf{r})$ is proportional to

$$p(s \mid \mathbf{r}) \propto \prod_{i}^{n} \frac{e^{-f_{\mathrm{i}}(s)} \cdot f_{\mathrm{i}}(s)^{r_{\mathrm{i}}}}{r_{\mathrm{i}}!} \cdot p(s) \tag{7.2}$$

where $f_{\mathrm{i}}(s)$ is the tuning curve of neuron *i* and r_{i} its spiking rate (Ma, Beck, Latham, & Pouget, 2006; Sanger, 1996). Using such a Bayesian decoder, very sharply tuned probability estimates can be inferred from a population of broadly tuned cells (see Figure 7.2, panel C).

Such a coding scheme was also described as gain encoding because the mean and the variance of the posterior distribution mainly depend on the gain of the noisy population code. This is due to the Poisson noise pattern where the variance of neural activity is directly proportional to its gain. Thus, the peak of activity can be treated as neural code for the posterior distribution with its position coding the mean and its gain coding the variance of the probability distribution. In this way, Poisson noise might be regarded as being beneficial instead of detrimental because it allows for implementing Bayesian inferences that require an estimate of the variance of encoded values (Knill & Pouget, 2004, p. 716 f.).

Taken together, probabilistic information might be represented as a distributed pattern of neuronal activity forming a population code. But how are these patterns processed to allow for a Bayesian combination of several stimulus attributes? Similar to the MLE model that is described in section 2.4, the neural activity of several population codes that have to be combined might be multiplied within an output layer consisting of a comparable number of cells (Pouget et al., 2003, p. 405 f.). This population code now has the same properties as an MLE integrator, thus its peak activity and variance are sensitive to the probability distributions in the input layers. Moreover, the multiplicative neural circuitries that are required for this kind of processing seem to be biologically plausible (Salinas & Abbott, 1996).

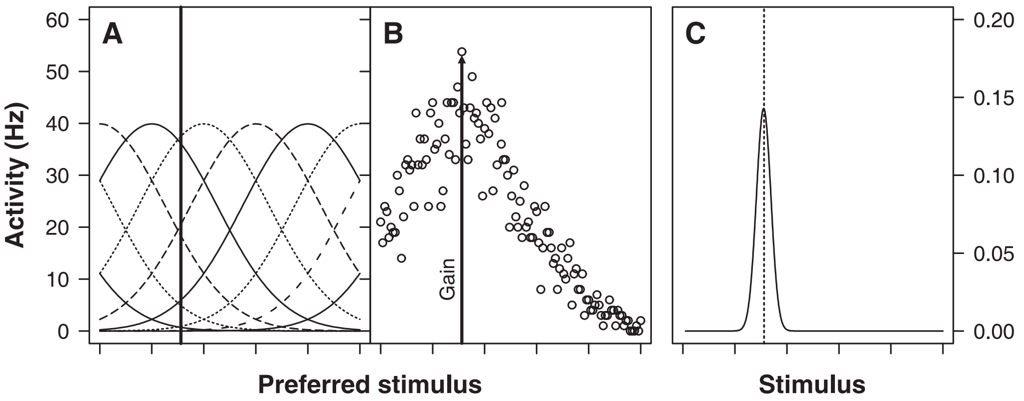


Figure 7.2. Illustration of a Bayesian computation on the neural level. Panel A shows idealized Gaussian tuning curves of several cells with different preferred stimuli. Panel B depicts the response pattern of more than 100 neurons with similar tuning curves to the stimulus that is shown as a vertical line in panel A. The cells have been ranked according to their preferred stimuli and the response pattern has been corrupted by independent Poisson noise. Panel C shows the posterior probability distribution resulting from the application of a Bayesian decoder (Sanger, 1996). The peak of this distribution (marked by the dotted line) closely corresponds the original stimulus in panel A.

A different kind of neural integration model was proposed by Deneve, Latham, and Pouget (2001). They constructed a network architecture that relies on the gain encoding mechanism that was described above to allow for Bayesian inferences. In this network, neural input layers are interconnected with a multidimensional intermediate layer which in turn projects to an output layer. Importantly, all connections within this network are bidirectional, thus, multisensory activities are fed back from the intermediate layer to the input cells. This network performs as a Bayesian integrator that is sensitive to single-cue reliabilities. When the input layers are initialized with noisy population codes representing several encoded cues, the network stabilizes after a few iterations onto smooth noiseless population codes in all layers. This evolved activity closely resembles the maximum likelihood estimate (cf., Deneve, Latham, & Pouget, 1999). Deneve and Pouget (2004, p. 256) constructed such a network to integrate visual and haptic size information and compared its output to the empirical data of Ernst and Banks (2002). Both input layers were corrupted by Poisson noise and the gains were adjusted to match the unimodal discrimination performance of human observers. Afterwards, the network's output in a bimodal condition was calculated and it turned out that single-cue weights as well as bimodal discrimination thresholds closely matched the empirically observed values. Thus, this kind of neural network which is based on a biologically plausible model allows for Bayesian calculations (cf., Yang & Zemel, 2000).

Taken together, simulated neural networks currently account for two main features of Bayesian approaches: the representation of several input channels as probabilistic information as well as their integration in a reliability-sensitive manner (cf., Witten & Knudsen, 2005). These models can also account for an influence of spatial proximity on cue integration by assuming that the weight of synaptic connections decreases as a function of distance between receptive fields. Moreover, the gain of specific cells is defined as the number of spikes in a brief interval of time (Sanger, 1996). Thus, the integration of several input layers depends on the moment of their initialization. This mechanism might explain the behaviorally relevant influence of temporal proximity on sensory integration.

What is currently not accounted for by neural models of sensory integration is the impact of cue consistency on the combined estimate. However, because the neural model of Deneve et al. (2001) includes feedback connections from a multisensory layer to the unisensory inputs, it might allow for estimating the consistency of several input layers. In turn, such information might be used to reweight the influence of discrepant input populations on the output layer. Whether or not such neural computations are possible within this kind of network remains to be addressed by future research.

7.4 Physiology of sensory integration

Although computational neuroscience has developed precise models of sensory integration that account for a number of behaviorally relevant parameters, the neurophysiological evidence for these schemes is generally weak. This is mostly due to the lack of methods for measuring neural population activity in vivo (Knill & Pouget, 2004). Nevertheless, single-cell recordings as well as noninvasive psychophysiological studies provided some evidence on the mechanisms of sensory integration within certain brain regions (Meredith, 2002). The pioneer work in this area focused on multisensory integration in the cat superior colliculus and is summarized in an influential book of Stein and Meredith (1993).

7.4.1 Multimodal integration in single neurons

The superior colliculus is a structure in the mammalian midbrain that consists of several layers. Whereas cells in the upper layers solely receive visual input, a substantial number of multisensory neurons can be identified in deeper layers of this structure. These cells respond to stimuli from more than one modality or their response pattern is altered by the presence of a stimulus from a second modality. Bimodal (e.g., audiovisual) as well as trimodal cells which respond to visual, auditory, and somatosensory stimulation have been identified in several species, for example cats (Meredith & Stein, 1986a), guinea pigs (King & Palmer, 1985), and primates (Wallace, Wilkinson, & Stein, 1996). Since most of these multisensory neurons project to premotor and motor areas of the brainstem and spinal cord (Meredith & Stein, 1985; Meredith, Wallace, & Stein, 1992), they have been thought to be involved in stimulus detection and orienting behavior.

Multisensory cells in the superior colliculus typically have large receptive fields which show a high degree of overlap between sensory modalities. Thus, different sensory channels converge on individual multisensory neurons that are topographically organized. Their modality specific receptive fields are in rough spatial register with one another and stimulating the organism with spatially corresponding multisensory signals typically leads to response enhancement in the multimodal cell. The proportional amount of response enhancement *RE* is usually calculated according to the following equation

$$RE = \frac{CM - SM_{\max}}{SM_{\max}} \cdot 100\% \tag{7.3}$$

with *CM* representing the mean number of spikes evoked by the combined stimulation in several modalities and SM_{max} the mean number of impulses evoked by the most effective single-modality stimulus. When presenting spatially discrepant signals to the organism, the response enhancement turns into response depression, thus, the stimulus that is additionally presented in a second modality reduces the response of the multisensory neuron as compared to its original single-modality response (Meredith & Stein, 1986b, 1996). A comparable pattern of results was also obtained for neurons in the anterior ectosylvian sulcus in the cat's cortex (Wallace, Meredith, & Stein, 1992).

In addition to these spatial factors, temporal disparity among combinations of different sensory stimuli was identified as a major determinant of multisensory integration in superior colliculus neurons. Although substantial response enhancements were even found for temporal discrepancies of 200 ms and more, the magnitude of enhancement generally decayed monotonically as temporal discrepancy between both stimuli increased (Meredith, Nemitz, & Stein, 1987).

Besides spatial and temporal constraints that affect multisensory integration in superior colliculus neurons, response enhancement was found to be largest when individual stimuli were comparably weak. Under these conditions, more than 1000% of response enhancement was reported for specific cells (Meredith & Stein, 1986a). Thus, the integration of sensory inputs was highly nonlinear and superadditive as the neural response to the combined stimulus exceeded the sum of responses to either stimulus component. When increasing stimulus intensity, the amount of enhancement typically decreases until reaching additive or even subadditive values. This integration rule that depends on the intensity of single stimuli was termed “inverse effectiveness” (cf. Stanford, Quessy, & Stein, 2005) and could be successfully modeled within a Bayesian framework (Anastasio, Patton, & Belkacem-Boussaid, 2000).

It is always problematic to translate results from single-cell recordings in anesthetized cats or primates to the overt behavior of animals or even humans in multisensory environments, but several efforts have been made to bridge this gap. As the superior colliculus seems to be specifically involved in orienting behavior, early studies tried to construct experimental situations to examine the impact of spatial alignment of audiovisual stimuli on multisensory orienting. In a study by Stein, Huneycutt, and Meredith (1988), one group of cats was trained to move quickly towards brief visual or auditory stimuli that were presented individually. Afterwards, both stimuli were presented simultaneously from the same spatial position. A second group of animals learned to approach visual stimuli only, thus, they had to ignore the auditory stimulus that was additionally presented on some occasions but originated from spatially discrepant positions. After finishing this paradigm, both groups of animals were retrained and accomplished the conditions of the other group, respectively. It turned out that cats were significantly better in approaching the combined stimulus when both signals were spatially consistent (response enhancement) and their accuracy dropped sharply when visual and auditory signals were presented from discrepant spatial locations (response depression). Thus, the overt behavioral response pattern closely resembled the observed spiking activity of superior colliculus neurons (Meredith & Stein, 1986b, 1996) which may indicate a functional link between neural activation patterns and behavioral activity. This orienting behavior did also closely correspond to the predictions of a Bayesian model that takes into account single-cue reliability (cf., Knill & Saunders, 2004) as well as prior preferences regarding the degree of cue coupling and the distribution of audiovisual target locations (Rowland, Stanford, & Stein, 2007).

Comparable studies were also conducted using human observers who had to redirect their gaze towards a multimodal target. For example, Frens, Van Opstal, and Van der Willigen (1995) showed that saccadic eye movements towards a visual target were accelerated when presenting a spatially and temporally aligned auditory signal. The amount of response enhancement decreased monotonically as the spatial (or temporal) discrepancy between both signals increased (cf., Harrington & Peck, 1998). A second study, also relying on the latency of saccadic eye movements, replicated these results and additionally provided evidence for an "inverse effectiveness" rule in human audiovisual integration (Corneil, Van Wanrooij, Munoz, & Van Opstal, 2002). In this study, the degree of response enhancement was inversely related to the auditory signal to noise ratio. Thus, saccadic response times were shorter in the bimodal condition as compared to either unimodal condition only when the reliability of the auditory signal was low. As its signal to noise ratio increased, the amount of response enhancement decreased. These results are in close agreement to properties of multimodal neurons in the superior colliculus and in both studies, the authors speculated about the involvement of this brain structure in the respective experimental task (Corneil et al., 2002, p. 449 ff.; Frens et al., 1995, p. 814 f.).

The mutimodal properties of superior colliculus neurons also resemble the behavioral results of numerous studies on the redundant signal effect (e.g., Miller, 1982, 1991). This effect describes the acceleration of response times to a multimodal signal in comparison to the reaction to either unimodal component. This effect could not solely be explained by statistical facilitation (i.e., responding to the product of the faster of two unimodal processing channels, Raab, 1962) and therefore exhibits certain characteristics of superadditivity. Moreover, it was shown that the principle of "inverse effectiveness" does also hold for this effect (Diederich & Colonius, 2004).

Taken together, single cells in the mammalian midbrain reveal a response pattern that closely resembles the results of behavioral studies with respect to the impact of stimulus intensity as well as spatial and temporal factors on multisensory integration (Holmes & Spence, 2005). Neurons with comparable properties were additionally found in many cortical regions extending into areas that were previously thought to be unisensory (Ghazanfar & Schroeder, 2006) and it seems reasonable to assume that these neural populations are involved in diverse multisensory tasks. Whereas these cells' properties are compatible with behavioral data regarding spatial and temporal constraints of multimodal integration (see section 7.1.2), they do not seem to account for a reliability sensitive integration of simultaneously available signals. Instead, the relative activity of these neurons to multimodal stimuli decreases monotonically as the spiking rate to single stimuli increases (Stanford et al., 2005). By contrast, the MLE-model which specifies a reliability-sensitive integration scheme predicts a stable gain in multimodal response enhancement as the signal to noise ratio of single-cue estimates increases (Gepshtein et al., 2005, p. 1014 f.). This apparent discrepancy might be resolved by assuming that individual signals are only integrated with respect to their reliability when they do not fully define the property that has to be estimated. Thus, when one cue unambiguously allows for interpreting a current scene, it may fully trigger behavioral responses with little or no influence of additional cues that are available. By contrast, when multiple cues are individually weak, the organism might benefit from their combination according to a reliability sensitive integration scheme. Alternatively, it might be supposed that simple, binary judgments (e.g., target detection) follow an "inverse effectiveness" rule whereas more complex, quantitative judgments (e.g., depth estimation) rely on an integration of individual signals that is sensible to cue reliabilities.

Some evidence for these different "modes" of integration was reported by Frens et al. (1995). Observers in this study were instructed to direct their gaze to visual targets as fast as possible while ignoring simultaneously presented auditory nontargets. When using signals with high intensity, the observers showed very straight saccadic eye movements towards the visual signal even when auditory signals were presented at different locations. In this case, only the saccadic latency was influenced by the auditory signal. However, when reducing the reliability of both signals, the saccade trajectories were less accurate. When visual and auditory signals were spatially discrepant in this condition, the saccades started in a direction that was between the two stimuli. Thus, the saccadic target seemed to represent a conglomerate of both single cues as predicted by a weighted additive integration of both signals. Assuming multiple "modes" of intersensory integration also allows accounting for empirical demonstrations of cue vetoing (e.g., Bülthoff & Mallot, 1988; O'Brien & Johnston, 2000; Rock & Victor, 1964; Singer & Day, 1969) and prior preferences of specific cues (e.g., Battaglia et al., 2003).

7.4.2 Multimodal integration in larger brain areas

Besides studies demonstrating multisensory convergence on single neurons in several cortical and subcortical regions (Ghazanfar & Schroeder, 2006; Stein & Meredith, 1993), recent research additionally focused on the identification of larger multisensory integration sites within the human brain by using hemodynamic (e.g., positron emission tomography, PET; functional magnetic resonance imaging, fMRI) or electromagnetic techniques (e.g., electroencephalography, EEG; magnetoencephalography, MEG; Calvert & Thesen, 2004). Unfortunately, the criteria for the identification of multisensory neurons cannot be directly transferred to these noninvasive psychophysiological techniques because the acquired signal relies on the composite activity of large neuronal ensembles. Thus, it is well possible to find clusters in the human brain that show comparable activity to stimuli from more than one modality or exhibit enhanced responding to bimodal stimuli as compared to the unimodal response pattern without including any multimodal cell (Laurienti, Perrault, Stanford, Wallace, & Stein, 2005, p. 290 f.). Thus, different criteria or experimental designs have to be developed to identify multisensory integration sites using these methods.

One such method was proposed by Calvert, Hansen, Iversen, and Brammer (2001). These authors used simple visual and auditory signals in an fMRI study to reveal audiovisual integration sites in the human brain. Participants passively attended unimodal as well as bimodal stimulus presentations. In the audiovisual condition, both signals were presented in a synchronous or asynchronous manner. In the statistical analyses, it was supposed that multisensory regions show similar characteristics as single cells in the superior colliculus (see section 7.4.1; Meredith, 2002). Thus, an area was defined to integrate audiovisual signals when its activity in the congruent audiovisual condition exceeded the sum of both unimodal responses (superadditive response enhancement). Moreover, the activity in the incongruent bimodal condition was supposed to be smaller than the largest unimodal response in this area (response suppression). Using these criteria, a network of brain areas was revealed that seemed to be involved in crossmodal audiovisual processing. This network included the superior colliculi as well as multiple regions in the temporal lobe, the frontal cortex, the parietal cortex, and the insula (q.v., Bushara, Grafman, & Hallett, 2001).

In line with these results, several regions in the human and primate cortex were repeatedly found to exhibit multisensory characteristics. These areas include regions in the parietal cortex (e.g., ventral and lateral intraparietal area), temporal cortex (e.g., zones within the superior temporal sulcus), and frontal cortex (e.g., ventral premotor cortex, ventrolateral prefrontal cortex, principal sulcus), as well as in the insula (for reviews see Calvert, 2001, Ghazanfar & Schroeder, 2006). However, it became also evident that the network of cortical regions engaged in multisensory tasks seems to largely depend on the particular combination of modalities, the characteristics of the information provided within each sensory channel, and the experimental paradigm used to identify multisensory integration sites (Calvert, 2001).

Besides crossmodal influences on neural activity in “higher” cortical areas, evidence has also accumulated that multisensory signals are capable of modulating activity in early sensory cortices that were traditionally considered as sensory-specific (for a review see Macaluso, 2006). For example, Macaluso, Frith, and Driver (2000) demonstrated that spatially congruent tactile stimulation amplified activity to visual signals in the contralateral lingual gyrus. By contrast, touch alone did not activate this region in the occipital lobe. These results were interpreted as relying on back-projections from multimodal parietal areas to the lingual gyrus. According to this reasoning, sensory signals are initially processed by sensory specific areas that subsequently converge on multimodal cells in higher regions. In turn, these multimodal sites are supposed to modulate the activity of sensory specific regions by using feed-back connections (cf., Driver &

Spence, 2000; Shimojo & Shams, 2001).

Comparable response enhancements in areas of the caudal auditory belt were recently reported when stimulating anesthetized monkeys with temporally aligned audio-tactile signals (Kayser, Petkov, Augath, & Logothetis, 2005). When presenting asynchronous auditory and tactile signals, the amount of response enhancement in the auditory cortex decreased sharply. In an additional experimental condition, the intensity of the auditory stimulus was decreased. As a result, a smaller volume in the auditory cortex was activated by bimodal stimuli. However, these voxels obeyed the law of inverse effectiveness (cf., Stanford et al., 2005). Thus, the amount of response enhancement increased as the effectiveness of the auditory stimulus decreased. Comparable results were reported for the visual modulation of activity in the auditory cortex of alert and anesthetized animals (Kayser, Petkov, Augath, & Logothetis, 2007).

Due to the fact that these multisensory interactions were observed in early regions of the processing hierarchy in anesthetized animals, it was speculated that they might rely on direct interconnections between sensory-specific areas instead of feedback connections from higher multimodal brain sites (Kayser et al., 2005, p. 380 f.). This reasoning was further supported by electrophysiological studies on the redundant signal effect. By comparing event-related brain potentials (ERPs) to audiovisual stimuli with the sum of unimodal ERPs, Molholm and colleagues (2002) revealed audiovisual interactions as early as 45 to 80 ms after stimulus onset. Even when taking into account potentially confounding common brain activity in unimodal and bimodal trails (cf., Teder-Salejarvi, McDonald, Di Russo, & Hillyard, 2002), crossmodal interactions at around 80 ms after stimulus onset were observed (Gondan & Röder, 2006). Such early intersensory interactions are extremely unlikely to be based on feedback projections from multimodal brain sites (Foxe & Schroeder, 2005).

The behavioral relevance of early intersensory interactions was further substantiated by psychophysiological studies on the so-called sound-induced illusory flash phenomenon (Shams et al., 2000). In this paradigm, observers frequently perceive multiple flashes when a single flash is accompanied by multiple auditory beeps. This effect was shown to be highly robust with respect to experimental parameters and stimulus modalities and seems to reflect a general mechanism of intersensory event perception (Shams et al., 2000, q.v., section 2.4.1). Using this paradigm, Shams, Kamitani, Thompson, and Shimojo (2001) revealed that auditory stimuli modulated visually evoked ERPs as early as 170 ms after stimulus onset. Moreover, ERPs were qualitatively similar between illusionary perceived and real flashes. Illusory perception of multiple flashes was accompanied by higher oscillatory and induced gamma band responses over the occipital cortex than veridically perceived trials (Bhattacharya, Shams, & Shimojo, 2002). Using MEG within this paradigm, it was subsequently demonstrated that sounds are capable of modulating the activity of occipital and parietal areas very early after stimulus onset (35-65 ms; Shams, Iwaki, Chawla, & Bhattacharya, 2005).

Taken together, these results indicate that intersensory perception is no strict feed-forward process from unisensory to multisensory sites. Instead, feed-back projections from higher multimodal sites (Macaluso et al., 2000) as well as direct interconnections between unisensory cortices seem plausible (Foxe & Schroeder, 2005; Schroeder & Foxe, 2005). However, the functional role of these connections has still to be determined. Recently, it was suggested that direct interconnections between sensory-specific cortices might serve to initiate unspecific preparatory mechanisms (possibly related to arousal) that increase stimulus processing efficiency. By contrast, feed-back projections from higher multimodal sites seem to be more specific with respect to spatial and temporal constraints of multisensory signals and may thus be more relevant for an integration of specific stimulus properties (Macaluso & Driver, 2005).

These studies reveal important insights into the process of multisensory integration in the brain and begin to explore how real-world constraints such as spatial and temporal proximity are represented at the neural level. However, sensory integration seems to additionally depend on the reliability and consistency of several signals and until now it remains unclear on how the brain differentiates between correlated signals that have to be combined and discrepant signals that have to be processed separately. A recent study on audiovisual integration provided first evidence for a gatekeeper function of the thalamus in this context (Baier, Kleinschmidt, & Müller, 2006). Participants had to accomplish a simple visual or auditory classification task in this study while brain activity was measured using fMRI. On each trial, a visual cue indicated the task-relevant channel and participants had to classify the target that appeared several seconds later. The target was always presented bimodally with correlated visual and auditory signals in one experimental condition and randomly paired signals in another condition that was accomplished on a different day. Interestingly, the neural activation in sensory-specific cortices differed between these conditions in the preparatory period before the target appeared. When signals were paired at random, activity was enhanced in the task-relevant system and suppressed in the task-irrelevant channel. Conversely, when stimuli were reliably associated, the cue induced increased activity in both sensory systems. Additionally, thalamic structures showed enhanced activity in the first as compared to the latter condition which may reflect a selection or gatekeeper process of the thalamus in this context. Taken together, this study demonstrated that the modulation of cortical activity depended on whether multimodal signals were assumed to be related or unrelated. Thus, it provides first evidence on neural mechanisms of sensory integration with respect to cue consistency.

7.4.3 Sensory integration of visual cues

Whereas much research is available on the physiological basis of multisensory integration, only few studies are available that focused on the integration of several visual cues. Recently, such intramodal integration was examined in the superior colliculus (Alvarado, Vaughan, Stanford, & Stein, 2007) which originally served as a model of multisensory integration in numerous studies (see section 7.4.1). Because individual neurons in deep layers of the superior colliculus have large receptive fields (Wallace et al., 1996), it was possible to stimulate them simultaneously with either two stimuli from the same (vision) or from different modalities (vision and audition). It turned out that the neurons showed very different response patterns in these two conditions. Whereas large degrees of response enhancement were found for multisensory stimulation, the simultaneous presentation of two visual signals typically resulted in small or absent effects of enhancement. For intramodal integration, neurons seemed to operate according to a maximum or an averaging rule. Thus, the activity for a simultaneous stimulation with two visual cues resembled the maximum activity of both stimulus components or corresponded to their average, respectively. These different modes for intersensory and intrasensory integration in the superior colliculus were further substantiated by a recent study of Alvarado, Stanford, Vaughan, and Stein (2007) which demonstrated that cortical projections into the superior colliculus (cf., Wallace & Stein, 1994) seem to be responsible for multisensory response enhancement but have no effect on the integration of several signals within one modality. It seems that two visual stimuli are represented as competitors whereas multimodal stimuli are processed as synergists within the superior colliculus. Thus, the former result in subadditive interactions whereas the latter give rise to superadditive response enhancement. Although these studies reveal important insights into unimodal processing in the superior colliculus, they are not representative for the typical case of intrasensory integration where two (or more) cues representing different aspects of a complex stimulus have to be combined to allow for veridical interpretations of a given scene. This is the case in visual depth perception where different cues are available and have to be combined to allow for precise depth estimating (see section 2.1.2).

Only few psychophysiological studies are available on visual depth perception and they mainly focused on the neural representation of single depth cues instead of their combination. An exception is a recent study by Welchman, Deubelius, Conrad, Bülthoff, and Kourtzi, (2005). These authors used the blood oxygen level dependent (BOLD) contrast as measured by fMRI to examine the integration of binocular disparity and perspective cues in the visual cortex. In early visual areas (V1, V2, V3, and V3a), the BOLD response was sensitive to changes in single cues irrespective of the combined percept. Thus, individual cues seemed to be processed separately in these areas. By contrast, higher visual areas in the ventral (lateral occipital complex, LOC) and the dorsal pathway (hMT+/V5) appeared to represent the perceived global shape. These areas were activated to a similar degree when metamers were presented that were psychophysically indistinguishable but consisted of slightly different disparity and perspective cues. However, because the BOLD response relies on the composite activity of large neuronal ensembles, these results do not necessarily imply that single cues are integrated into a global 3-D representation in these areas. A comparable response pattern could also result from the simultaneous activity of several subpopulations of unimodal neurons in the respective brain regions (q.v., Laurienti et al., 2005, p. 291). To differentiate between these two possibilities, single-cell recordings are necessary. Such studies examining cue integration in the monkey cortex were recently published (for a summary of these studies see Tsutsui, Taira, & Sakata, 2005).

One set of studies focused on the caudal part of the lateral intraparietal sulcus which belongs to the dorsal processing stream. In this region, neurons have been identified that selectively respond to 3-D surface orientation. Interestingly, many of these cells showed similar response selectivity to surfaces that were defined by perspective cues or stereoscopic disparity (Tsutsui, Jiang, Yara, Sakata, & Taira, 2001). When both cues were simultaneously available, the neural response magnitude resembled the summation of unimodal spiking activities. It may thus be concluded that both cues were integrated in these cells. Comparable results have also been reported for the integration of texture gradients and stereoscopic disparity in the caudal intraparietal sulcus (Tsutsui, Sakata, Naganuma, & Taira, 2002).

A second target region for the integration of depth cues was identified in the anterior part of the inferior temporal cortex. In the lower bank of the superior temporal sulcus which belongs to the ventral processing stream, neurons have been found that selectively respond to stereoscopically defined 3-D shapes (Janssen, Vogels, & Orban, 2000). More recently, it was demonstrated that some proportion of these neurons are also sensitive to other depth cues (Liu, Vogels, &

Orban, 2004). In this study, rhesus monkeys passively viewed tilted planes that were either defined by texture cues or by a disparity gradient (random dot stereograms). Approximately half of the neurons that were examined responded selectively to specific tilts of the stimulus irrespective of the cue type. Moreover, different texture patterns led to a similar spiking activity. Thus, these cells' activity seems to correspond to an integrated estimate of 3-D shape.

These studies indicate that individual depth cues are integrated in certain regions of the ventral and dorsal processing streams to form a 3-D representation of the currently observed scene. Moreover, this integration seems to occur automatically as was demonstrated by Sereno, Trinath, Augath, and Logothetis (2002) with anesthetized monkeys. However, these studies reveal little insight into the physiological process of intrasensory integration with respect to single-cue properties (e.g., cue reliability; see section 7.1.1) and multi-cue interrelations (e.g., cue consistency; see section 7.1.3). It has still to be determined how different degrees of cue reliability are represented on the neural level and how they affect cue integration. Moreover, although some evidence for an additive integration of neural activity was reported (Tsutsui et al., 2001, p. 2864), it remains to be seen whether this computation holds for different degrees of cue reliability and consistency.

7.5 Conclusions and suggestions for future research

Taken together, the present study provided evidence for an influence of cue consistency on the integration of visual depth cues that is not accounted for by currently popular cue integration schemes (MLE-models, Ernst & Bülthoff, 2004). In several experiments, single-cue weights were shown to depend on cue consistency in addition to single-cue reliability (Experiments 1, 2, and 4). A Bayesian model that takes into account these pairwise consistencies in addition to single-cue reliability (see section 2.5) provided a better fit to the empirical data in comparison to a model disregarding covariations among cues (Experiment 1 and 2). This influence of cue consistency could also be obtained when using a multiple-observation task (cf., Experiment 3) that helps to avoid the potentially problematic assessment of single-cue reliabilities. However, in Experiment 4, it was shown that cue interactions can be highly specific. Whereas shading and texture as well as texture and disparity seemed to be processed independently, consistent shading and disparity cues led to a downweighting of the dissenting texture cue without reducing the discriminability of the stimulus at hand. This response pattern cannot be explained by the MLE model and suggests that at least with respect to some cues, consistency plays an important role. This specificity of cue interactions allows to account for former experimental data demonstrating an MLE-like integration of texture and disparity cues (Hillis et al., 2004; Knill & Saunders, 2003) and extends results from cross-modal studies that showed a consistency-based reweighting of visual and haptic cues after repeated exposure to stimuli with artificial cue correlations (Atkins et al., 2001, 2003; Ernst et al., 2000). Furthermore, this study sheds light on basic principles of robust fusion (Landy et al., 1995) that have been largely disregarded by previous research. The ability of sensory systems to provide relatively stable estimates of physical properties in complex and variable situations seems to depend on cue consistency in addition to spatial and temporal proximity as well as cue reliability.

The Bayesian model that was outlined in section 2.5 allows deriving several additional hypotheses that could be tested by future psychophysical studies. First, the reliability of the combined estimate should be sensitive to the amount of cue conflicts (q.v., Figure 7.1, panel A). This could be tested rather easily by varying the degree of cue discrepancy while measuring the observer's sensitivity. Second, cues are assumed to remain separately accessible after taking consistencies into account. At first glance, this hypothesis seems to conflict with an empirical study showing that visual cues are completely fused during sensory integration (Hillis et al., 2002). However, in yet another study, it was shown that single-cue information remains accessible after cue combination (Hogervorst & Brenner, 2004). Additionally, this hypothesis is substantiated by a recent study from our laboratory demonstrating that selective attention is capable of enhancing cue weights after taking cue reliability and consistency into account (Marian, 2007). These data further support the extended Bayesian integration scheme by demonstrating that individual cues are not entirely fused in visual depth perception.

The current study demonstrated an influence of consistency on the integration of visual depth cues but it remains questionable whether such an influence also occurs in multimodal integration. Earlier studies demonstrated that the coupling of several cues might be acquired during extensive exposure to such correlations (Atkins et al., 2001, 2003; Ernst et al., 2000; q.v., Ernst, 2007) but until now, it is unclear whether such intermodal consistencies are also detected and processed on a trial-by-trial basis as demonstrated for the case of intramodal integration in the current study. Such an examination could be realized by using multimodal event perception as experimental paradigm (e.g., Shams et al., 2000, 2002). It was already shown that visual and auditory (Andersen et al., 2004; Shams et al., 2005), auditory and tactile (Bresciani et al., 2005; Bresciani & Ernst, 2007), and visual and tactile signals (Bresciani et al., 2006) are integrated in a reliability dependent manner in these tasks. Using a comparable but trimodal setup with visual, auditory, and tactile signals, it could be determined whether these cues are integrated into a combined estimate solely on the basis of their relative reliabilities or whether pairwise consistencies additionally affect the number of perceived events.

Current computational models do not account for an influence of cue consistency on the process of sensory integration and it has yet to be determined how such computations might be implemented in neural networks. Recent models (Deneve et al., 2001) already include feedback connections between a multimodal layer and the unimodal inputs that seem to be in line with physiological evidence from multisensory integration (Driver & Spence, 2000; Macaluso, 2006). However, it remains unclear whether these connections are relevant for detecting cue consistencies and whether the influence of cue interactions on sensory integration can be adequately captured by these computational models.

To date, only few psychophysiological studies have examined neural mechanisms of sensory integration in visual depth perception. These studies provided evidence for a convergence of depth information from several cues on neurons in the dorsal (Tsutsui et al., 2005) and ventral stream of the visual system (Liu et al., 2004). However, the neural representation and processing of various factors that are known to affect visual depth perception (cf., section 7.1) have not been systematically examined in these studies. Future research might concentrate on how differential reliability and consistency of visual depth cues is represented at the neural level. Single-cell animal studies as well as research using noninvasive psychophysiological methods might be utilized to achieve this aim. Especially the multiple observation task that was successfully used in Experiment 3 and 4 of the current study may be suitable for such research because it allows for quantifying influences of reliability and consistency while avoiding potentially problematic single-cue conditions. Thus, the physiology of cue integration with respect to these factors may be examined by using perceptually comparable stimuli. Moreover, Experiment 4 provided evidence for highly specific cue interactions in visual depth perception. By measuring ERPs in a comparable setup, it should be possible to determine electrophysiological differences between simple additive integration rules and non-linear combination schemes (e.g., consistent shading and disparity cues). Moreover, the role of gamma band responses in cue combination might be assessed in this setup. Such high-frequency synchronization of neural activity seems to be important for visual feature integration in object perception (Singer & Gray, 1995) and may also be involved in sensory integration during visual depth perception.

To sum up, the current study demonstrated that the scope of contemporary models of sensory integration which strongly focus on single-cue reliability might be too narrow. Interrelations among cues with respect to spatial and temporal proximity but also regarding the consistency of estimates have substantial impact on cue integration. Thus, the perceptual system seems to be optimized for robust perception in an ecologically valid environment. It remains a challenging task for future research to integrate knowledge from psychophysical studies, computational neuroscience, and physiological research into a global framework of intrasensory and multimodal integration that accounts for these principles.

8. References

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

Adams, W. J., & Mamassian, P. (2004). Bayesian combination of ambiguous shape cues. *Journal of Vision*, *4*(10):7, 921-929.

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

Alvarado, J. C., Stanford, T. R., Vaughan, J. W., & Stein, B. E. (2007). Cortex mediates multisensory but not unisensory integration in superior colliculus. *Journal of Neuroscience*, *27*, 12775-12786.

Alvarado, J. C., Vaughan, J. W., Stanford, T. R., & Stein, B. E. (2007). Multisensory versus unisensory integration: contrasting modes in the superior colliculus. *Journal of Neurophysiology*, *97*, 3193-3205.

Anastasio, T. J., Patton, P. E., & Belkacem-Boussaid, K. (2000). Using Bayes' rule to model multisensory enhancement in the superior colliculus. *Neural Computation*, *12*, 1165-1187.

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, *21*, 301-308.

Andersen, T. S., Tiippana, K., & Sams, M. (2005). Maximum Likelihood Integration of rapid flashes and beeps. *Neuroscience Letters*, *380*, 155-160.

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

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

Backus, B. T., Banks, M. S., Ee, R. van, & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, *39*, 1143-1170.

Baier, B., Kleinschmidt, A., & Müller, N. G. (2006). Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. *Journal of Neuroscience*, *26*, 12260-12265.

Banks, M. S. (2004). Neuroscience: what you see and hear is what you get. *Current Biology*, *14*, R236- R238.

Banks, M. S., Hooge, I. T., & Backus, B. T. (2001). Perceiving slant about a horizontal axis from stereopsis. *Journal of Vision*, *1*(2):1, 55-79.

Barlow, H., & Földiák, P. (1989). Adaptation and decorrelation in the cortex. In C. Miall, R. M. Durbin, & G. J. Mitchison (Eds.), *The computing neuron* (pp. 54-72). Wokingham: Addison-Wesley.

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

Beers, R. J. van, Sittig, A. C., & Denier van der Gon, J. J. (1996). How humans combine simultaneous proprioceptive and visual position information. *Experimental Brain Research*, *111*, 253-261.

Beers, R. J. van, Sittig, A. C., & Denier van der Gon, J. J. (1998). The precision of proprioceptive position sense. *Experimental Brain Research*, *122*, 367-377.

Beers, R. J. van, Sittig, A. C., & Denier van der Gon, J. J. (1999). Integration of proprioceptive and visual position information: an experimentally supported model. *Journal of Neurophysiology*, *81*, 1355-1364.

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

Berg, B. G. (1989). Analysis of weights in multiple observation tasks. *Journal of the Acoustical Society of America*, *86*, 1743-1746.

Berg, B. G. (1990). Observer efficiency and weights in a multiple observation task. *Journal of the Acoustical Society of America*, *88*, 149-158.

Berkeley, G. (1732). *An essay towards a new theory of vision* (4th ed.) [Electronic version]. Retrieved January 21, 2008, from http://www.maths.tcd.ie/~dwilkins/Berkeley/Vision/1732B/Vision.pdf

Bermant, R. I., & Welch, R. B. (1976). Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. *Perceptual and Motor Skills*, *43*, 487-493.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin and Review*, *5*, 482-489.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*, 578-584.

Bhattacharya, J., Shams, L., & Shimojo, S. (2002). Sound-induced illusory flash perception: role of gamma band responses. *Neuroreport*, *13*, 1727-1730.

Blake, A., Bülthoff, H. H., & Sheinberg, D. (1993). Shape from texture: ideal observers and human psychophysics. *Vision Research*, *33*, 1723-1737.

Bradshaw, M. F., Parton, A. D., & Glennerster, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Research*, *40*, 3725-3734.

Bradshaw, M. F., & Rogers, B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Research*, *36*, 3457-3468.

Bresciani, J., 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.

Bresciani, J., & Ernst, M. O. (2007). Signal reliability modulates auditory-tactile integration for event counting. *Neuroreport*, *18*, 1157-1161.

Bresciani, J., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., & Kheddar, A. (2005). Feeling what you hear: auditory signals can modulate tactile tap perception. *Experimental Brain Research*, *162*, 172-180.

Bruno, N., & Cutting, J. E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General*, *117*, 161-170.

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research*, *33*, 919-933.

Buckley, D., Frisby, J. P., & Blake, A. (1996). Does the human visual system implement an ideal observer theory of slant from texture? *Vision Research*, *36*, 1163-1176.

Bülthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: stereo and shading. *Journal of the Optical Society of America A: Optics, Image Science & Vision*, *5*, 1749-1758.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *Journal of Neuroscience*, *21*, 300-304.

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, *11*, 1110-1123.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage*, *14*, 427-438.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology - Paris*, *98*, 191-205.

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

Cohen, J. D. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Corneil, B. D., Van Wanrooij, M., Munoz, D. P., & Van Opstal, A. J. (2002). Auditory-visual interactions subserving goal-directed saccades in a complex scene. *Journal of Neurophysiology*, *88*, 438-454.

Cumming, B. G., Johnston, E. B., & Parker, A. J. (1993). Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Research*, *33*, 827-338.

Curran, W., & Johnston, A. (1996). The effect of illuminant position on perceived curvature. *Vision Research*, *36*, 1399-1410.

Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General*, *113*, 198-216.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience.* Cambridge, MA: MIT Press.

Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience*, *2*, 740-745.

Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, *4*, 826-831.

Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology - Paris*, *98*, 249-258.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*, 1388-1404.

Doorschot, P. C., Kappers, A. M., & Koenderink, J. J. (2001). The combined influence of binocular disparity and shading on pictorial shape. *Perception & Psychophysics*, *63*, 1038-1047.

Dosher, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, *26*, 973-990.

Drewing, K., & Ernst, M. O. (2006). Integration of force and position cues for shape perception through active touch. *Brain Research*, *1078*, 92-100.

Driver, J., & Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Current Biology*, *10*, R731-R735.

Erens, R. G., Kappers, A. M., & Koenderink, J. J. (1993). Perception of local shape from shading. *Perception & Psychophysics*, *54*, 145-156.

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. In G. Knoblich, M. Grosjean, I. Thornton, & M. Shiffrar (Eds.), *Perception of the human body from the inside out* (pp. 105-131). New York: Oxford University Press.

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

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

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

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162-169.

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, *63*, 719-725.

Fine, I., & Jacobs, R. A. (1999). Modeling the combination of motion, stereo, and vergence angles cues to visual depth. *Neural Computation*, *11*, 1297-1330.

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport*, *16*, 419-423.

Frens, M. A., Van Opstal, A. J., & Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception & Psychophysics*, *57*, 802-816.

Frisby, J. P., Buckley, D., Wishart, K. A., Porrill, J., Garding, J., & Mayhew, J. E. (1995). Interaction of stereo and texture cues in the perception of three-dimensional steps. *Vision Research*, *35*, 1463-1472.

Georgeson, M. A., & Schofield, A. J. (2002). Shading and texture: separate information channels with a common adaptation mechanism? *Spatial Vision*, *16*, 59-76.

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., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision*, *5*(11):7, 1013-1023.

Ghahramani, Z., Wolpert, D. M., & Jordan, M. I. (1997). Computational Models of Sensorimotor Integration. In P. Morasso, & V. Sanguineti (Eds.), *Self-Organization, Computational Maps and Motor Control* (pp. 117-147). Amsterdam: Elsevier.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*, 278-285.

Goldstein, E. B. (2007). *Sensation and Perception* (7th ed.). Belmont, CA: Thomson Wadsworth.

Gondan, M., Lange, K., Rösler, F., & Röder, B. (2004). The redundant target effect is affected by modality switch costs. *Psychonomic Bulletin & Review*, *11*, 307-313.

Gondan, M., & Röder, B. (2006). A new method for detecting interactions between the senses in event-related potentials. *Brain Research*, *1073-1074*, 389-397.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Greenwald, H. S., Knill, D. C., & Saunders, J. A. (2005). Integrating visual cues for motor control: A matter of time. *Vision Research*, *45*, 1975-1989.

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 (PNAS)*, *103*, 483-488.

Hampel, F. R. (1974). The Influence Curve and Its Role in Robust Estimation. *Journal of the American Statistical Association*, *69*, 383-393.

Harrington, L. K., & Peck, C. K. (1998). Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Experimental Brain Research*, *122*, 247-252.

Hay, J. C., Pick, H. L., & Ikeda, K. (1965). Visual capture produced by prism spectacles. *Psychonomic Science*, *2*, 215-216.

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research*, *179*, 595-606.

Heron, J., Whitaker, D., & McGraw, P. V. (2004). Sensory uncertainty governs the extent of audio-visual interaction. *Vision Research*, *44*, 2875-2884.

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. 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.

Hirsh, I. J., & Sherrick Jr., C. E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, *62*, 423-432.

Hogervorst, M. A., & Brenner, E. (2004). Combining cues while avoiding perceptual conflicts. *Perception*, *33*, 1155-1172.

Holmes, N. P., & Spence, C. (2005). Multisensory integration: space, time and superadditivity. *Current Biology*, *15*, R762-R764.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*, 106-154.

Huynh, H., & Feldt, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, *1*, 69-82.

Ichikawa, M., Saida, S., Osa, A., & Munechika, K. (2003). Integration of binocular disparity and monocular cues at near threshold level. *Vision Research*, *43*, 2439-2449.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the 'ventriloquism' effect. *Perceptual and Motor Skills*, *37*, 967-979.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*, 3621-3629.

Jacobs, R. A. (2002a). Visual cue integration for depth perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic Models of the Brain* (pp. 61-76). Cambridge, MA: Bradford Book.

Jacobs, R. A. (2002b). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*, 345-350.

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

Jäkel, F., & Ernst, M. O. (2003). Learning to combine arbitrary signals from vision and touch. In I. Oakley, S. O. Modhrain, & F. Newell (Eds.), *Eurohaptics 2003 Conference Proceedings* (pp. 276-290). Dublin: Trinity College Dublin & Media Lab Europe.

Janssen, P., Vogels, R., & Orban, G. A. (2000). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science*, *288*, 2054-2056.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, *31*, 1351-1360.

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research*, *34*, 2259-2275.

Johnston, E. B., Cumming, B. G., & Parker, A. J. (1993). Integration of depth modules: stereopsis and texture. *Vision Research*, *33*, 813-826.

Jones, B., & O'Neil, S. (1985). Combining vision and touch in texture perception. *Perception & Psychophysics*, *37*, 66-72.

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex. *Neuron*, *48*, 373-384.

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2007). Functional imaging reveals visual modulation of specific fields in auditory cortex. *Journal of Neuroscience*, *27*, 1824-1835.

Kinchla, R. A. (1974). Detecting target elements in multielement arrays: a confusability model. *Perception & Psychophysics*, *15*, 149-158.

King, A. J., & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research*, *60*, 492-500.

Knill, D. C. (1998a). Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vision Research*, *38*, 1655-1682.

Knill, D. C. (1998b). Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Research*, *38*, 1683-1711.

Knill, D. C. (1998c). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research*, *38*, 2635-2656.

Knill, D. C. (2005). Reaching for visual cues to depth: the brain combines depth cues differently for motor control and perception. *Journal of Vision*, *5*(2):2, 103-115.

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.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*, 712-719.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge, MA: 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*, 2539-2558.

Knill, D. C., & Saunders, J. A. (2004). *Bayesian models of sensory cue integration* [Electronic version]. Retrieved January 21, 2008, from http://www.irp.oist.jp/ocnc/2004/public/Knill.pdf

Koenderink, J. J., & Doorn, A. J. van (1995). Relief: pictorial and otherwise. *Image and Vision Computing*, *13*, 321-334.

Koenderink, J. J., Doorn, A. J. van, Christou, C., & Lappin, I. S. (1996). Perturbation study of shading in pictures. *Perception*, *25*, 1009-1026.

Koenderink, J. J., Doorn, A. J. van, & Kappers, A. M. (1992). Surface perception in pictures. *Perception & Psychophysics*, *52*, 487-496.

Köhler, W. (1947). *Gestalt Psychology.* New York: Liveright Publishing Corporation.

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 & Vision*, *18*, 2307-2320.

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.

Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., & Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research*, *166*, 289-297.

Lederman, S. J., & Abbott, S. G. (1981). Texture perception: studies of intersensory organization using a discrepancy paradigm, and visual versus tactual psychophysics. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 902-915.

Lederman, S. J., Thorne, G., & Jones, B. (1986). Perception of texture by vision and touch: Multidimensionality and intersensory integration. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 169-180.

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, *63*, 1279-1292.

Liu, B., & Todd, J. T. (2004). Perceptual biases in the interpretation of 3D shape from shading. *Vision Research*, *44*, 2135-2145.

Liu, Y., Vogels, R., & Orban, G. A. (2004). Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex. *Journal of Neuroscience*, *24*, 3795-3800.

Lutfi, R. A. (1995). Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks. *Journal of the Acoustical Society of America*, *97*, 1333-1334.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432-1438.

Macaluso, E. (2006). Multisensory processing in sensory-specific cortical areas. *Neuroscientist*, *12*, 327-338.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends in Neurosciences*, *28*, 264-271.

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, *289*, 1206-1208.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, *24*, 253-257.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Maloney, L. T., & Landy, M. S. (1989). A statistical framework for robust fusion of depth information. In W. A. Pearlman (Ed.), *Visual Communications and Image Processing IV, Proceedings of the SPIE*, *1199*, 1154-1163.

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, MA: Bradford Book.

Marian, H. (2007). *Die Rolle der situativen Relevanz bei der sensorischen Integration visueller Tiefenhinweise [The role of task relevance during sensory integration of visual depth cues]*. Johannes Gutenberg-Universität Mainz: Unpublished Diploma Thesis.

Maronna, R. A., Martin, D., & Yohai, V. J. (2006). *Robust statistics*. Chichester: Wiley.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.

Meinhardt, G., Persike, M., Mesenholl, B., & Hagemann, C. (2006). Cue combination in a combined feature contrast detection and figure identification task. *Vision Research*, *46*, 3977-3993.

Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Cognitive Brain Research*, *14*, 31-40.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience*, *7*, 3215-3229.

Meredith, M. A., & Stein, B. E. (1985). Descending efferents from the superior colliculus relay integrated multisensory information. *Science*, *227*, 657-659.

Meredith, M. A., & Stein, B. E. (1986a). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*, 640-662.

Meredith, M. A., & Stein, B. E. (1986b). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, *365*, 350-354.

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*, *75*, 1843-1857.

Meredith, M. A., Wallace, M. T., & Stein, B. E. (1992). Visual, auditory and somatosensory convergence in output neurons of the cat superior colliculus: multisensory properties of the tecto-reticulo-spinal projection. *Experimental Brain Research*, *88*, 181-186.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247-279.

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 160-169.

Mingolla, E., & Todd, J. T. (1986). Perception of solid shape from shading. *Biological Cybernetics*, *53*, 137-151.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research*, *14*, 115-128.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, *17*, 154-163.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, *257*, 1357-1363.

Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research*, *25*, 499-507.

O'Brien, J., & Johnston, A. (2000). When texture takes precedence over motion in depth perception. *Perception*, *29*, 437-452.

O'Leary, A., & Wallach, H. (1980). Familiar size and linear perspective as distance cues in stereoscopic depth constancy. *Perception & Psychophysics*, *27*, 131-135.

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

Palmer, S. E. (1999). *Vision science: photons to phenomenology*. Cambridge, MA: MIT Press.

Parker, A. J., Cumming, B. G., Johnston, E. B., & Hurlbert, A. C. (1995). Multiple cues for threedimensional shape. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 351–364). Cambridge, MA: MIT Press.

Pentland, A. (1989). Shape information from shading: a theory about human perception. *Spatial Vision*, *4*, 165-182.

Perotti, V. J., Todd, J. T., Lappin, J. S., & Phillips, F. (1998). The perception of surface curvature from optical motion. *Perception & Psychophysics*, *60*, 377-388.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge & Keagan Paul.

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgements of spatial direction. *Perception & Psychophysics*, *6*, 203-205.

Pöppel, E., & Harvey, L. O. J. (1973). Light-difference threshold and subjective brightness in the periphery of the visual field. *Psychologische Forschung*, *36*, 145-161.

Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, *26*, 381-410.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574-590.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111-163.

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 Royal Society of London, Series B: Biological Sciences*, *273*, 2159-2168.

Rock, I., & Victor, J. (1964). Vision and touch: an experimentally created conflict between the two senses. *Science*, *143*, 594-596.

Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception*, *24*, 155-179.

Rosas, P., Wagemans, J., Ernst, M. O., & Wichmann, F. A. (2005). Texture and haptic cues in slant discrimination: reliability-based cue weighting without statistically optimal cue combination. *Journal of the Optical Society of America A: Optics, Image Science & Vision*, *22*, 801-809.

Rosas, P., Wichmann, F. A., & Wagemans, J. (2004). Some observations on the effects of slant and texture type on slant-from-texture. *Vision Research*, *44*, 1511-1135.

Rosas, P., Wichmann, F. A., & Wagemans, J. (2007). Texture and object motion in slant discrimination: failure of reliability-based weighting of cues may be evidence for strong fusion. *Journal of Vision*, *7*(6):3, 1-21.

Rowland, B., Stanford, T., & Stein, B. (2007). A Bayesian model unifies multisensory spatial localization with the physiological properties of the superior colliculus. *Experimental Brain Research*, *180*, 153-161.

Säfström, D., & Edin, B. B. (2004). Task requirements influence sensory integration during grasping in humans. *Learning & Memory*, *11*, 356-363.

Salinas, E., & Abbott, L. F. (1996). A model of multiplicative neural responses in parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, *93*, 11956-11961.

Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. *Journal of Neurophysiology*, *76*, 2790-2793.

Saunders, J. A., & Backus, B. T. (2006). Perception of surface slant from oriented textures. *Journal of Vision*, *6*(9):3, 882-897.

Schofield, A. J., & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: separate mechanisms with similar behaviour. *Vision Research*, *39*, 2697-2716.

Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. *Current Opinion in Neurobiology*, *15*, 454-458.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.

Sereno, M. E., Trinath, T., Augath, M., & Logothetis, N. K. (2002). Three-dimensional shape representation in monkey cortex. *Neuron*, *33*, 635-652.

Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005). Early modulation of visual cortex by sound: an MEG study. *Neuroscience Letters*, *378*, 76-81.

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

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, *14*, 147-152.

Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport*, *12*, 3849-3852.

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

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, *11*, 505-509.

Singer, G., & Day, R. H. (1969). Visual capture of haptically judged depth. *Perception & Psychophysics*, *5*, 315-316.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, *18*, 555-586.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*, 7-10.

Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *Journal of Neuroscience*, *25*, 6499-6508.

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. (1988). Neurons and behavior: the same rules of multisensory integration apply. *Brain Research*, *448*, 355-358.

Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses*. Cambridge, MA: MIT Press.

Stevens, K. A. (1981). The information content of texture gradients. *Biological Cybernetics*, *42*, 95-105.

Teder-Salejarvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research*, *14*, 106-114.

Thurlow, W. R., & Jack, C. E. (1973). Certain determinants of the 'ventriloquism effect'. *Perceptual and Motor Skills*, *36*, 1171-1184.

Thurlow, W. R., & Rosenthal, T. M. (1976). Further study of existence regions for the 'ventriloquism effect'. *Journal of the American Audiology Society*, *1*, 280-286.

Tittle, J. S., Norman, J. F., Perotti, V. J., & Phillips, F. (1998). The perception of scale-dependent and scale-independent surface structure from binocular disparity, texture, and shading. *Perception*, *27*, 147-166.

Todd, J. T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences*, *8*, 115-121.

Todd, J. T., & Akerstrom, R. A. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 242-255.

Todd, J. T., & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 583-595.

Todd, J. T., Norman, J. F., Koenderink, J. J., & Kappers, A. M. (1997). Effects of texture, illumination, and surface reflectance on stereoscopic shape perception. *Perception*, *26*, 807-822.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, *23*, 775-785.

Triesch, J., Ballard, D. H., & Jacobs, R. A. (2002). Fast temporal dynamics of visual cue integration. *Perception*, *31*, 421-434.

Tsutsui, K., Jiang, M., Yara, K., Sakata, H., & Taira, M. (2001). Integration of perspective and disparity cues in surface-orientation-selective neurons of area CIP. *Journal of Neurophysiology*, *86*, 2856-2867.

Tsutsui, K., Sakata, H., Naganuma, T., & Taira, M. (2002). Neural correlates for perception of 3D surface orientation from texture gradient. *Science*, *298*, 409-412.

Tsutsui, K., Taira, M., & Sakata, H. (2005). Neural mechanisms of three-dimensional vision. *Neuroscience Research*, *51*, 221-229.

Violentyev, A., Shimojo, S., & Shams, L. (2005). Touch-induced visual illusion. *Neuroreport*, *16*, 1107-1110.

Vuong, Q. C., Domini, F., & Caudek, C. (2006). Disparity and shading cues cooperate for surface interpolation. *Perception*, *35*, 145-155.

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research*, *91*, 484-488.

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

Wallace, M. T., & Stein, B. E. (1994). Cross-modal synthesis in the midbrain depends on input from cortex. *Journal of Neurophysiology*, *71*, 429-432.

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology*, *76*, 1246-1266.

Warren, D. H., & Cleaves, W. T. (1971). Visual-proprioceptive interaction under large amounts of conflict. *Journal of Experimental Psychology*, *90*, 206-214.

Welchman, A. E., Deubelius, A., Conrad, V., Bülthoff, H. H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience*, *8*, 820-827.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. [Laws of Organization in Perceptual Forms]. *Psychological Research*, *4*, 301-350.

Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: Merging auditory and visual worlds. *Neuron*, *48*, 489-496.

Yang, Z., & Zemel, R. (2000). Managing uncertainty in cue combination. In S. A. Solla, T. K. Leen, & K.-R. Müller (Eds.), *Advances in Neural Information Processing Systems 12* (pp. 80-86). Cambridge, MA: MIT Press.

Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*, *33*, 2685-2696.

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

Appendix

Appendix A

1. Mathematical derivation of the posterior probability distribution parameters of the MLE model

As outlined in section 2.4, the MLE model is a special case of a Bayesian integration scheme where noises of individual estimates are independent and Gaussian. Furthermore, the prior is proportional to Kronecker's delta function, thus, cues are completely fused into an integrated estimate. Given these preconditions, the posterior probability density function corresponds to a multiplication of the Gaussian distributions of individual cue estimates (Oruç et al., 2003). Thus, for two cues with expected values $\hat{s}_1$ and $\hat{s}_2$, and corresponding standard deviations σ1 and σ2, the posterior probability density function is

$$p(S \mid s_1, s_2) = k \cdot e^{-\frac{1}{2}\left(\frac{\hat{s}_1 - s}{\sigma_1}\right)^2} \cdot e^{-\frac{1}{2}\left(\frac{\hat{s}_2 - s}{\sigma_2}\right)^2} \tag{A.1}$$

k is a normalization constant that ensures that the integral of the posterior probability distribution is equal to 1. Equation A.1 can be rewritten to

$$p(S \mid s_1, s_2) = k \cdot e^{-\frac{1}{2}\left(\left(\frac{\hat{s}_1 - s}{\sigma_1}\right)^2 + \left(\frac{\hat{s}_2 - s}{\sigma_2}\right)^2\right)} \tag{A.2}$$

This normalized product of two Gaussian distributions is again a normal distribution (Oruç et al., 2003, p. 2467) with the expected value $\hat{s}_c$ and the variance σ_c^2. The expected value $\hat{s}_c$ is the maximum of the continuous probability function A.2. It can be calculated by deriving A.2 by s which yields

$$p'(S \mid s_1, s_2) = p(S \mid s_1, s_2) \cdot \left(\frac{\hat{s}_1}{\sigma_1^2} + \frac{\hat{s}_2}{\sigma_2^2} - \frac{s}{\sigma_1^2} - \frac{s}{\sigma_2^2}\right) \tag{A.3}$$

The zero-crossing of this function reveals the expected value of the posterior probability distribution $\hat{s}_c$ which equals to

$$\hat{s}_c = \mathrm{E}[p(S \mid s_1, s_2)] = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \cdot \hat{s}_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \cdot \hat{s}_2 \tag{A.4}$$

The standard deviation of the posterior probability distribution σc equals the distance between the inflection point of equation A.2 and $\hat{s}_c$. A second derivation of A.2 by *s* yields

$$p''(S \mid s_1, s_2) = p'(S \mid s_1, s_2) \cdot \left(\left(\frac{\hat{s}_1}{\sigma_1^2} + \frac{\hat{s}_2}{\sigma_2^2} - \frac{s}{\sigma_1^2} - \frac{s}{\sigma_2^2} \right)^2 - \frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \right) \tag{A.5}$$

The zero-crossings of this function reveal the inflection points s_x lying at

$$s_x = \frac{\hat{s}_1\sigma_2^2 + \hat{s}_2\sigma_1^2 \pm \sigma_1\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1^2 + \sigma_2^2} \tag{A.6}$$

As outlined above, the standard deviation of the posterior probability density function σc is equal to the absolute difference between s_x and $\hat{s}_c$

$$\sigma_c = \sqrt{\mathrm{Var}[p(S \mid s_1, s_2)]} = |s_x - \hat{s}_c| \tag{A.7}$$

Inserting A.6 and A.4 into equation A.7 yields

$$\sigma_c = \left| \frac{\hat{s}_1\sigma_2^2 + \hat{s}_2\sigma_1^2 \pm \sigma_1\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1^2 + \sigma_2^2} - \frac{\hat{s}_1\sigma_2^2 + \hat{s}_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \right| \tag{A.8}$$

which can be simplified to

$$\sigma_c = \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{A.9}$$

Thus, when two cues with independent and Gaussian noises are fused into an integrated estimate using the above mentioned prior, its expected value $\hat{s}_c$ and its standard deviation σc can be calculated from the probability distributions of both single cues according to equation A.4 and A.9, respectively.

Appendix B

1. Discrepancies between the single- and multiple-cue conditions of Experiment 1

In Experiment 1, the reliability of each single cue was determined by using a depth of 300 pixels in the standard cylinder. The individual JNDs from this task were used to infer the cue weights according to the MLE framework. In the multiple-cue condition, however, cylinder depths of 200, 240, 280, and 320 pixels were used and it remains questionable whether the calculated single-cue weights can be adequately used in this condition or whether they might be supposed to change as a function of cylinder depth. For example, results from slant discrimination studies suggest that the reliability of texture cues increases with the slant angle of planar surfaces (Knill 1998b; Knill & Saunders, 2003; Rosas et al., 2004). The estimation of an elliptic hemi-cylinder's depth shares several features with a slant discrimination task and the question arises whether reliabilities of single cues might be supposed to change as a function of depth. This would pose a problem for the application of the MLE framework in Experiment 1 because cylinder depths differed between the within-cue discrimination task and the multiple-cue conditions.

I would like to put forward two arguments that help refuting this criticism: First, tangential angles to the cylinder's surface do not differ by more than 12° between the single-cue and the multiple-cue conditions of Experiment 1 across all possible fixation points (see derivation in section 1.1 below). Thus, large differences between the reliabilities of single cues as a function of depth are unlikely to occur. Second, a post-hoc study using one observer (MG) revealed that reliabilities of single cues were indeed relatively stable across different depths of the standard (see section 1.2 below).

1.1 Maximum difference between tangential angles

An elliptic surface can be regarded as a composition of an infinite number of slanted surfaces. The slant angle at each point of the ellipse corresponds to the angle α of the tangent at this point (see Figure B.1). The absolute angle at the fixation point (x_f, y_f) ranges from 0° to 90°, depending on the position of the line of sight x_f. Assuming that the center of the ellipse corresponds to the origin, this angle can be calculated by

$$\alpha = \arctan\left(\frac{b^2 x_\mathrm{f}}{a^2 y_\mathrm{f}}\right) \tag{B.1}$$

with

$$y_\mathrm{f} = b \cdot \sqrt{1 - \frac{x_f^2}{a^2}} \tag{B.2}$$

and a and b denoting the radii of the ellipse.

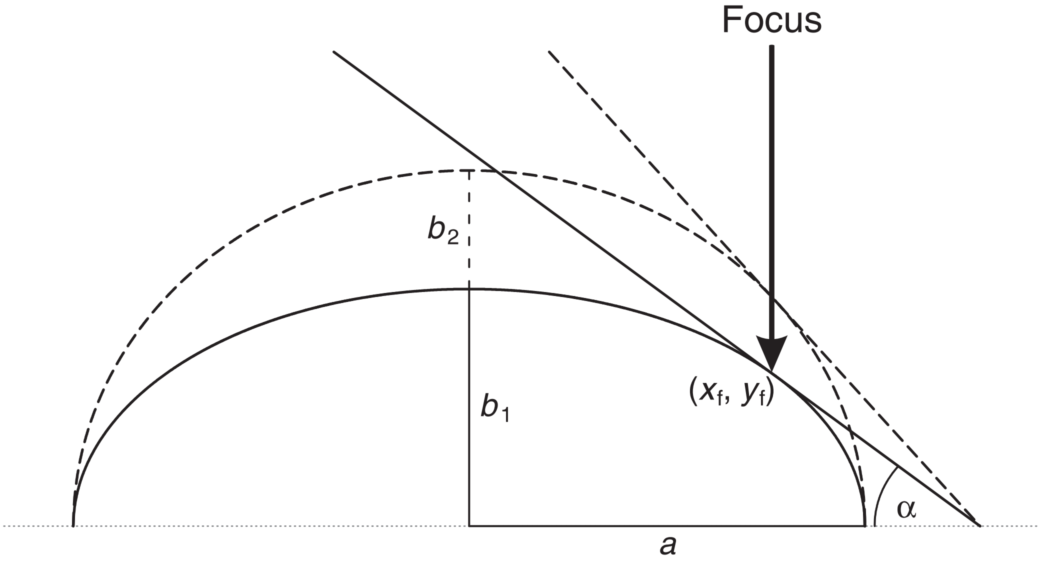


Figure B.1. Schematic drawing of two hemi-ellipses with an equal radius a and an unequal radius b. Additionally, tangents at a fixation point (x_f, y_f) are depicted.

Using equation B.1, the slant difference δ between two elliptic surfaces with different radii b can be calculated for each fixation point (x_f, y_f) by

$$\delta = \arctan\left(\frac{b_2^2 x_\mathrm{f}}{a^2 y_\mathrm{f}}\right) - \arctan\left(\frac{b_1^2 x_\mathrm{f}}{a^2 y_\mathrm{f}}\right) \qquad \text{(B.3)}$$

By substituting y_f with equation B.2 and a subsequent derivation by x_f, it can be shown that the difference between both slant angles is maximal at the line of sight

$$x_\mathrm{max} = \frac{a^2}{\sqrt{b_1 b_2 + a^2}} \qquad \text{(B.4)}$$

The maximum difference between both slants can be calculated by inserting x_max in equation B.3. This yields the maximum difference δ_max between the slant angles of the tangents to two elliptic surfaces across all possible fixation points:

$$\delta_\mathrm{max} = \arctan\left(\sqrt{\frac{b_1}{b_2}}\right) - \arctan\left(\sqrt{\frac{b_2}{b_1}}\right) \qquad \text{(B.5)}$$

In Experiment 1, the radius of the ellipse in the single-cue condition was $b_1 = 150$ pixels and in the multiple-cue condition it ranged from $b_2 = 160$ to 100 pixels. Inserting these values in equation B.5 yields maximum differences of $\delta_\mathrm{max} = -1.85°$ to $11.54°$ between the tangential angles to the cylinder's surface across all fixation points.

1.2 Empirical demonstration of roughly stable single-cue reliabilities across different depths

To test whether single-cue reliabilities vary as a function of cylinder depth, a post-hoc study was carried out. Observer MG accomplished a within-cue discrimination task that was similar to the procedure of Experiment 1 with the exception that the standard stimulus had a depth of 200, 240, 280, or 320 pixels. The JND for each cue at each depth of the standard stimulus was measured once in each of six experimental sessions with randomly interleaved tracks. All sessions were accomplished within a period of three weeks.

Figure B.2 depicts the JNDs of each cue as a function of cylinder depth. Additionally, pooled values that closely resemble the results of the original within-cue discrimination task using a standard depth of 300 pixels (see Figure 3.3) are shown on the right side of the figure. Obviously, JNDs remained relatively stable across different cylinder depths for shading and motion cues. With respect to the texture cue, JNDs tended to be largest for roughly circular cylinders although this effect was relatively weak (see error bars in Figure B.2). In slant discrimination tasks, the accuracy of slant from texture judgments typically varies to a much larger degree, for example by an order of magnitude from low to high slants (Knill & Saunders, 2003). Thus, texture cues might be differentially processed in slant discrimination and depth estimation tasks. The small differences in the JNDs of the texture cue as a function of cylinder depth were systematically taken into account in Experiment 2.

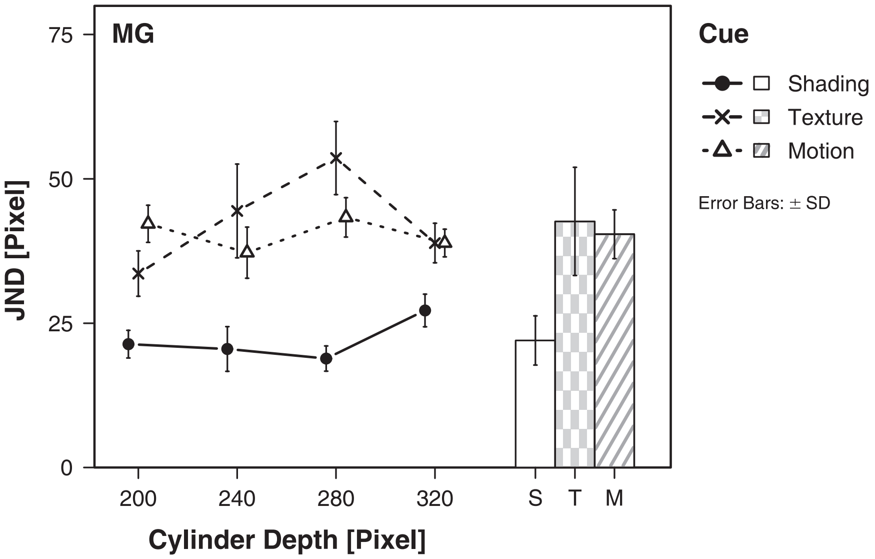


Figure B.2. Just noticeable differences (JND) from the within-cue discrimination tasks as a function of cue and depth of the standard cylinder for observer MG. Pooled values across all depths are depicted on the right side.

Appendix C

1. Single subject data of Experiment 4

In the statistical analyses of Experiment 4, only aggregated data across participants was reported (see 6.3). In this section, weights and *d'* values are depicted separately for each condition and observer to allow for an inspection of the generalizability of the statistical results to a single-subject level.

1.1 Comparison of the control condition to conditions with pairwise cue consistencies

The statistical analyses on the perceptual weights in section 6.3 revealed a significant condition by cue interaction that was primarily due to a decreased weight of the texture cue in the condition with consistent shading and disparity cues. As can be seen from Figure C.1, this response pattern was obtained for all observers except BK and JG. Reweighting single cues in the other conditions was less systematically, thus producing no significant differences from the weights in the control condition. Regarding the observer's sensitivity, reduced *d'* values in the condition with consistent texture and disparity cues were obtained for six of ten observers (see Table C.1). A similarly reduced *d'* when shading and texture indicated a comparable depth of the cylinder was observed for all participants except CR and MW. Thus, these results occurred consistently across observers.

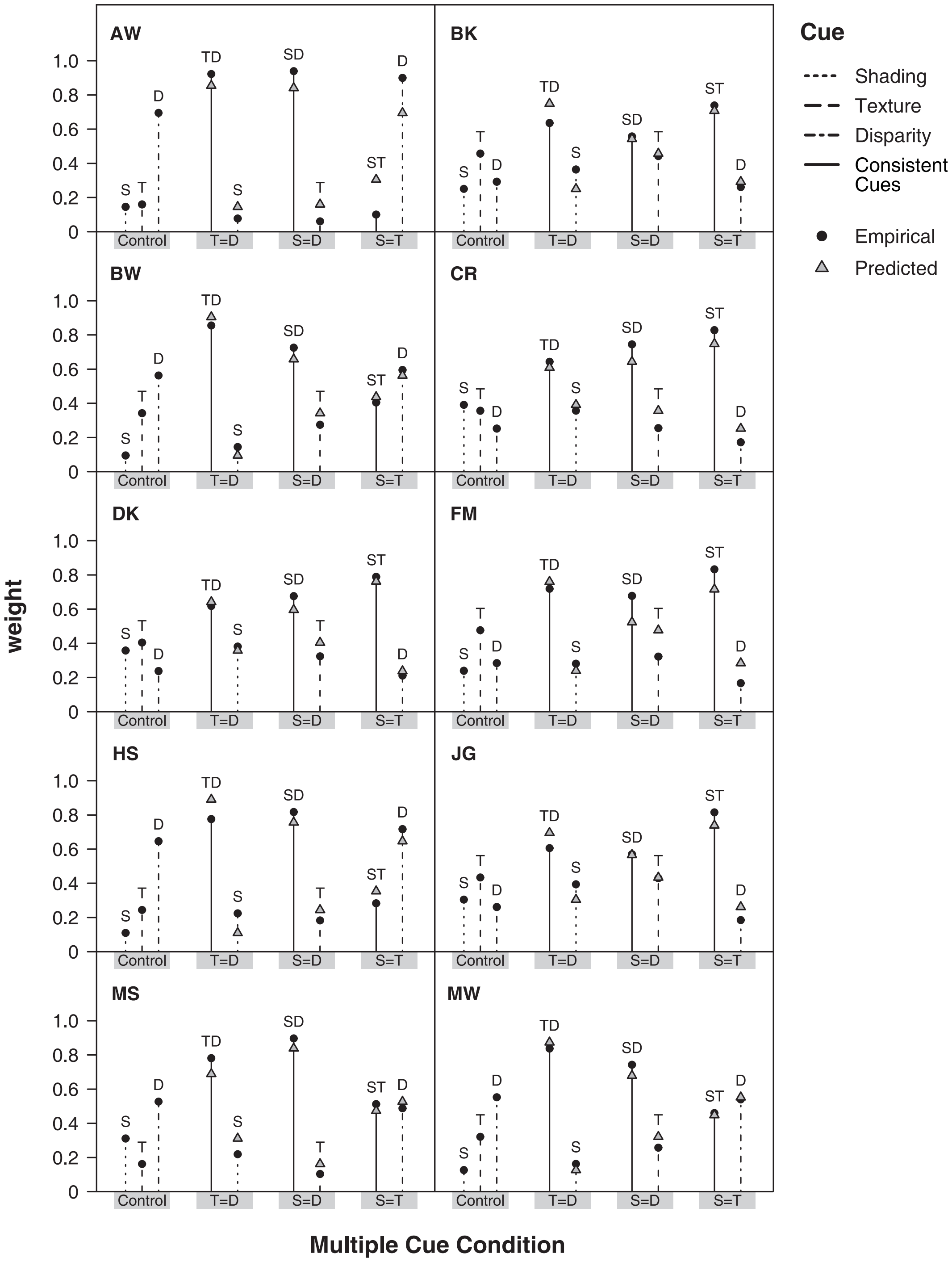


Figure C.1. Perceptual weights for each cue or cue combination as a function of experimental condition, depicted separately for each observer. The characters S, T and D denote the shading, texture and disparity cue, respectively. The black circles depict the empirical weights; gray triangles correspond to the predictions of the MLE model (see 6.3.2).

Table C.1. Empirical (*d'*$_e$) and predicted sensitivity measures (*d'*$_p$) for each observer in the control condition and the conditions with two consistent and one deviant depth cue.

		Deviant Cue					
	Control	Shading		Texture		Disparity	
Observer	d'_e	d'_e	d'_p	d'_e	d'_p	d'_e	d'_p
AW	1.04	1.05	0.93	1.11	0.94	0.89	1.01
BK	1.02	0.96	0.90	0.95	0.95	0.89	0.91
BW	1.04	0.89	0.88	1.01	0.99	0.92	1.01
CR	0.86	0.80	0.81	0.96	0.80	1.03	0.78
DK	1.25	1.00	1.10	1.18	1.11	1.00	1.04
FM	0.95	0.90	0.85	1.04	0.89	0.82	0.86
HS	1.02	1.01	0.89	1.07	0.96	0.75	1.00
JG	1.00	1.02	0.90	1.02	0.93	0.88	0.89
MS	1.11	0.90	1.01	0.89	0.93	1.07	1.04
MW	1.09	1.00	0.92	1.07	1.01	1.11	1.04
Across observers	1.04 (0.10)	0.95 (0.08)	0.92 (0.08)	1.03 (0.08)	0.95 (0.08)	0.94 (0.11)	0.96 (0.09)

Note. Average values across observers are depicted in the last row with the standard deviation in brackets.

1.2 Comparison of the control condition to conditions with reduced single-cue reliabilities

Decreasing the shading or texture reliability led to a downweighting of this cue for five of ten observers, respectively (see Figure C.2). More stable results were obtained for the disparity cue that was heavily downweighted by all observers when its reliability was reduced. Correspondingly, *d'* was not systematically affected by changes in the reliability of the shading or texture cue, respectively, but it was largely reduced when the reliability of the disparity cue decreased (see Table C.2).

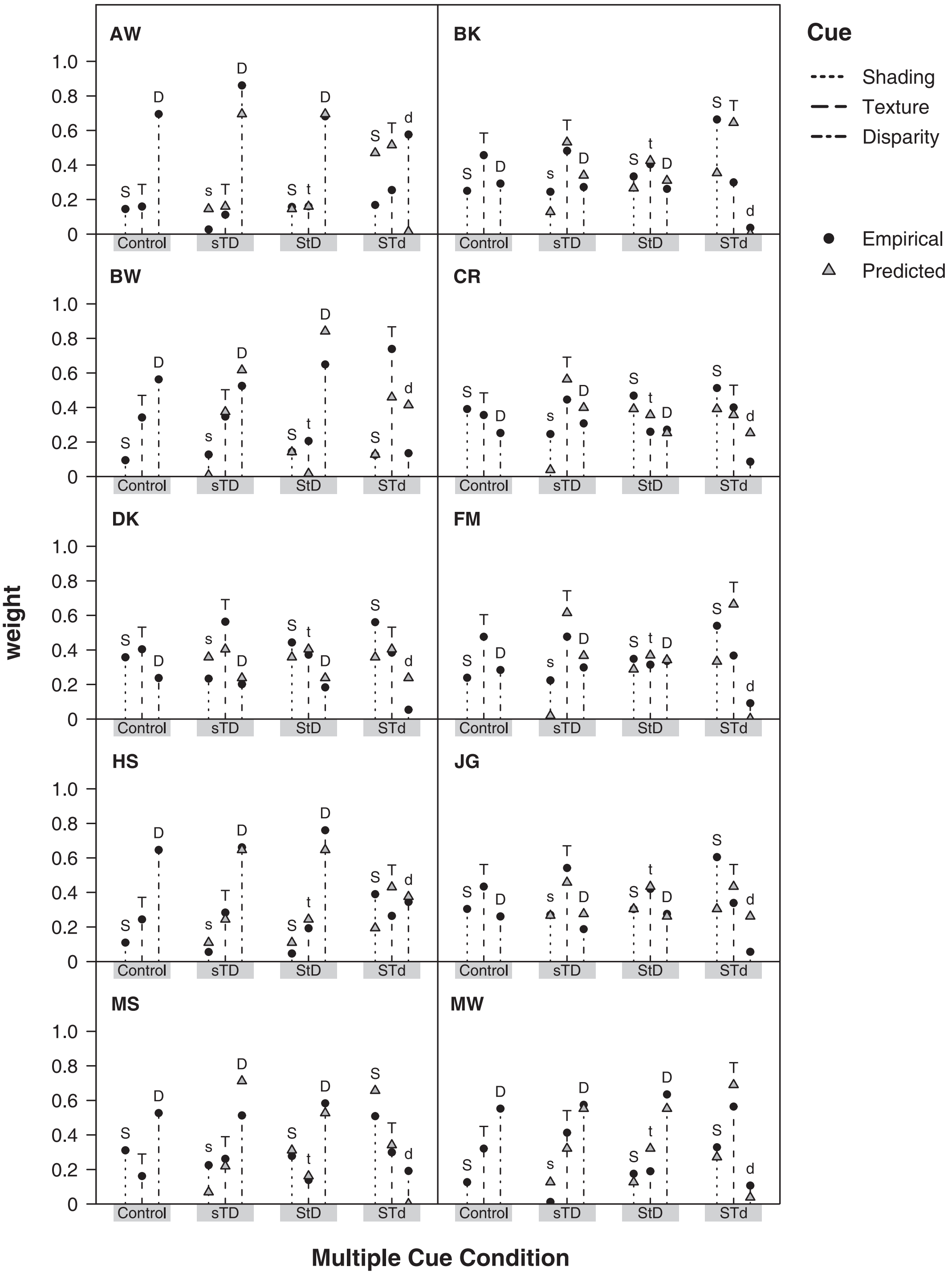


Figure C.1. Perceptual weights for each cue as a function of experimental condition, depicted separately for each observer. The characters S, T and D denote the shading, texture and disparity cue, respectively, and reduced single-cue reliabilities are indicated by lower case letters. The black circles depict the empirical weights; gray triangles correspond to the predictions of the MLE model (see 6.3.2).

Table C.2. Empirical (d'_e) and predicted sensitivity measures (d'_p) for each observer in the control condition and the conditions with one less reliable depth cue.

		Unreliable Cue					
	Control	Shading		Texture		Disparity	
Observer	d'_e	d'_e	d'_p	d'_e	d'_p	d'_e	d'_p
AW	1.04	1.17	1.04	1.15	1.04	0.53	0.75
BK	1.02	0.96	0.94	1.00	1.00	0.69	0.84
BW	1.04	0.91	0.98	0.85	0.83	0.91	0.98
CR	0.86	0.71	0.69	1.08	0.86	0.85	0.86
DK	1.25	1.29	1.25	1.31	1.25	1.34	1.25
FM	0.95	0.81	0.82	0.90	0.90	0.58	0.79
HS	1.02	1.10	1.02	1.14	1.02	0.93	0.92
JG	1.00	0.96	0.98	1.05	1.00	1.11	1.00
MS	1.11	1.02	0.94	1.09	1.11	0.24	0.84
MW	1.09	1.07	1.09	1.10	1.09	0.87	0.83
Across observers	1.04 (0.10)	1.00 (0.17)	0.98 (0.15)	1.07 (0.13)	1.01 (0.13)	0.80 (0.31)	0.91 (0.15)

Note. Average values across observers are depicted in the last row with the standard deviation in brackets.

Appendix D

Fragebogen Experiment 3   Code: Datum:

Bitte füllen Sie diesen Bogen wahrheitsgemäß und vollständig aus.

1. Alter: _______ Jahre   2. Geschlecht: ____________

3. Studienfach / Beruf: ____________________   4. Semesterzahl: __________

5. Höchster Schulabschluss:
○ Hauptschule
○ Realschule
○ Gymnasium
○ Fachhochschule
○ Universität
○ Anderer: _________________________

6. Haben Sie eine Sehschwäche?   ○ ja   ○ nein

7. Wenn ja, welche?
○ Farbsehschwäche / Farbenblindheit: __________
○ Kurzsichtigkeit: links _____ dpt / rechts _____ dpt
○ Weitsichtigkeit: links _____ dpt / rechts _____ dpt
○ Andere Fehlsichtigkeit: _____________________

8. Ist die Sehschwäche korrigiert (Brille, Kontaktlinsen, o.ä.)?   ○ ja   ○ nein

9. Was ist Ihr dominantes Auge?   ○ links ○ rechts   ○ keine Präferenz

10. An wie vielen psychologischen Experimenten haben Sie bisher teilgenommen?

○ keinen   ○ wenigen   ○ vielen   ○ sehr vielen

11. Wie sind Sie auf dieses Experiment aufmerksam geworden?

……………………………………………………………………………………
……………………………………………………………………………………

12. Hatten Sie Vorinformationen bzgl. dieses Experimentes?   ○ ja   ○ nein

Wenn ja, welche?   ………………………………………………………………
………………………………………………………………
………………………………………………………………

13. Was denken Sie, war der Zweck des Experiments?

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

14. Wie häufig haben Sie ein intuitives Tiefenurteil gefällt?

○ nie ○ manchmal ○ oft ○ immer

15. Haben Sie bei Ihren Antworten eine bestimmte Strategie verfolgt? ○ ja ○ nein

Wenn ja, welche?

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

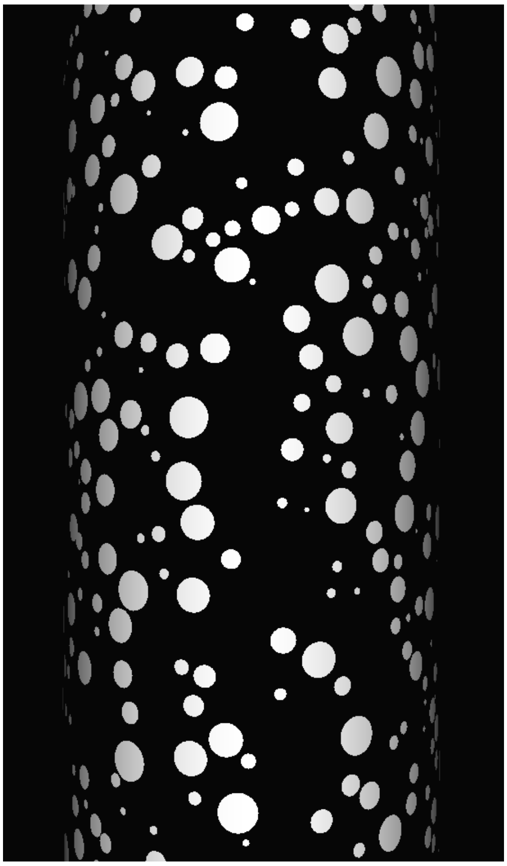
…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

16. Sie sehen hier das Bild eines statischen Zylinders aus dem Experiment. Kennzeichnen Sie bitte alle Hinweise, die zur Einschätzung der Zylindertiefe verwendet werden können und geben Sie an, welche Hinweise Sie verwendet haben.



Haben Sie noch andere Tiefenhinweise bemerkt / verwendet, die im obigen Bild nicht erkennbar sind?

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

17. Ist Ihnen an den gezeigten Stimuli etwas Besonderes aufgefallen? ○ ja ○ nein

Wenn ja, was?

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

18. Die Stimuli enthielten teilweise sich widersprechende Tiefeninformationen. Haben Sie das bemerkt? ○ ja ○ nein

Wenn ja, wann? Während der …

○ 1. Sitzung ○ 2. Sitzung ○ 3. Sitzung ○ 4. Sitzung

19. Fiel Ihnen die Tiefeneinschätzung insgesamt schwer?

○ ja, sehr ○ ja, eher schon ○ nein, eher nicht ○ nein, gar nicht

20. War das Experiment insgesamt anstrengend für Sie?

○ ja, sehr ○ ja, eher schon ○ nein, eher nicht ○ nein, gar nicht

21. Fiel Ihnen die Tiefeneinschätzung nach und nach leichter?

○ ja, sehr ○ ja, eher schon ○ nein, eher nicht ○ nein, gar nicht

22. War Ihnen das Experiment insgesamt unangenehm?

○ ja, sehr ○ ja, eher schon ○ nein, eher nicht ○ nein, gar nicht

Abschlusskritik

Zum Abschluss haben Sie nun noch die Gelegenheit **Anmerkungen, Kommentare, Kritik und Anregungen** zum Experiment anzugeben.
Sind Ihnen bestimmte Dinge sehr angenehm oder unangenehm aufgefallen?

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

→ Vielen Dank für die Teilnahme an dieser Untersuchung!!

Appendix E

Fragebogen Experiment 4 **Code:** **Datum:**

Bitte füllen Sie diesen Bogen wahrheitsgemäß und vollständig aus.

1. Alter: _______ Jahre 2. Geschlecht: ____________

3. Studienfach / Beruf: ____________________ 4. Semesterzahl: __________

5. Höchster Schulabschluss:
○ Hauptschule
○ Realschule
○ Gymnasium
○ Fachhochschule
○ Universität
○ Anderer: _________________________

6. Haben Sie eine Sehschwäche? ○ ja ○ nein

7. Wenn ja, welche?
○ Farbsehschwäche / Farbenblindheit: __________
○ Kurzsichtigkeit: links _____ dpt / rechts _____ dpt
○ Weitsichtigkeit: links _____ dpt / rechts _____ dpt
○ Andere Fehlsichtigkeit: _____________________

8. Ist die Sehschwäche korrigiert (Brille, Kontaktlinsen, o.ä.)? ○ ja ○ nein

9. Was ist Ihr dominantes Auge? ○ links ○ rechts ○ keine Präferenz

10. An wie vielen psychologischen Experimenten haben Sie bisher teilgenommen?

○ keinen ○ wenigen ○ vielen ○ sehr vielen

11. Wie sind Sie auf dieses Experiment aufmerksam geworden?

…………………………………………………………………………………
…………………………………………………………………………………

12. Hatten Sie Vorinformationen bzgl. dieses Experimentes? ○ ja ○ nein

Wenn ja, welche? ……………………………………………………………
……………………………………………………………
……………………………………………………………

13. Was denken Sie, war der Zweck des Experiments?

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

14. Wie häufig haben Sie ein intuitives Tiefenurteil gefällt?

○ nie ○ manchmal ○ oft ○ immer

15. Haben Sie bei Ihren Antworten bestimmte Strategien verfolgt? ○ ja ○ nein

16. Unterschieden sich diese Strategien, je nachdem welcher Stimulus gezeigt wurde? ○ ja ○ nein

17. Sie sehen nun auf der Powerwall einen typischen Stimulus, der im Experiment verwendet wurde. Beschreiben Sie bitte, welche Tiefenhinweise in diesem Stimulus enthalten sind und geben Sie dabei an, ob Sie diese Hinweise für Ihre Antworten verwendet haben

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

18. Ist Ihnen an den gezeigten Stimuli etwas Besonderes aufgefallen? o ja o nein

Wenn ja, was?

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

19. Sie sehen nun hintereinander verschiedene Stimuli, die im Experiment verwendet wurden. Geben Sie bitte jeweils an, wie schwer Ihnen die Einschätzung der Tiefe des Stimulus gefallen ist und welche Strategie Sie dabei verfolgten.

Stimulus 1:
Schwierigkeit der Tiefenschätzung:

o hoch   o eher hoch   o eher niedrig   o niedrig

Verwendete Strategie:

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

Stimulus 2:
Schwierigkeit der Tiefenschätzung:

o hoch   o eher hoch   o eher niedrig   o niedrig

Verwendete Strategie:

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

Stimulus 3:
Schwierigkeit der Tiefenschätzung:

o hoch   o eher hoch   o eher niedrig   o niedrig

Verwendete Strategie:

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

Stimulus 4:

Schwierigkeit der Tiefenschätzung:

o hoch o eher hoch o eher niedrig o niedrig

Verwendete Strategie:

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

Stimulus 5:

Schwierigkeit der Tiefenschätzung:

o hoch o eher hoch o eher niedrig o niedrig

Verwendete Strategie:

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

Stimulus 6:

Schwierigkeit der Tiefenschätzung:

o hoch o eher hoch o eher niedrig o niedrig

Verwendete Strategie:

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

Stimulus 7:

Schwierigkeit der Tiefenschätzung:

o hoch o eher hoch o eher niedrig o niedrig

Verwendete Strategie:

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

20. Wenn Sie das gesamte Experiment beurteilen, fiel Ihnen die Tiefeneinschätzung insgesamt schwer?

o ja, sehr o ja, eher schon o nein, eher nicht o nein, gar nicht

21. War das Experiment insgesamt anstrengend für Sie?

o ja, sehr o ja, eher schon o nein, eher nicht o nein, gar nicht

22. Fiel Ihnen die Tiefeneinschätzung nach und nach leichter?

o ja, sehr o ja, eher schon o nein, eher nicht o nein, gar nicht

23. War Ihnen das Experiment insgesamt unangenehm?

o ja, sehr o ja, eher schon o nein, eher nicht o nein, gar nicht

Abschlusskritik

Zum Abschluss haben Sie nun noch die Gelegenheit **Anmerkungen, Kommentare, Kritik und Anregungen** zum Experiment anzugeben.
Sind Ihnen bestimmte Dinge sehr angenehm oder unangenehm aufgefallen?

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

→ Vielen Dank für die Teilnahme an dieser Untersuchung!!

Zusammenfassung

Da die visuelle Wahrnehmung auf einer zweidimensionalen, retinalen Projektion der betrachteten Szenerie beruht, muss räumliche Tiefe aus verschiedenen Tiefenhinweisen erschlossen werden. Die Kombination dieser Merkmale führt nachfolgend zu einem stabilen Perzept. Aktuelle Modelle sensorischer Integration schreiben den Reliabilitäten einzelner Tiefenmerkmale eine prominente Rolle zu, vernachlässigen dabei jedoch deren Interaktionen untereinander. In der vorliegenden Studie wurde ein erweitertes Bayesianisches Modell erarbeitet und in vier Experimenten überprüft, das sowohl Reliabilität als auch Konsistenz verschiedener Wahrnehmungskanäle berücksichtigt. Probanden schätzten die räumliche Tiefe visuell präsentierter Halbzylinder mit elliptischer Grundfläche ein. In den Experimenten 1 und 2 wurden partiell konsistente Schattierungs-, Textur- und Bewegungsinformationen verwendet. Die gewonnenen empirischen Daten ließen sich gut durch das erweiterte Integrationsmodell erklären, was die Bedeutung der Konsistenz von Tiefenkriterien unterstützt. Um auf die problematische Messung der Reliabilitäten einzelner Tiefenhinweise verzichten zu können, wurde in Experiment 3 ein neueres Verfahren erfolgreich umgesetzt, mit dem sich perzeptuelle Gewichte einzelner Merkmale in komplexen Stimuli schätzen lassen. Dieses Verfahren wurde auch in Experiment 4 angewandt, um die Integration von stereoskopischer Disparität, Schattierungs- und Texturgradienten zu untersuchen. Die Reliabilität einzelner Tiefenkriterien wirkte sich proportional auf deren Gewichtung aus. Merkmalsinteraktionen spielten jedoch nur in Bezug auf konsistente Schattierungs- und Disparitätsinformation eine entscheidende Rolle. Andere Merkmalskombinationen ließen sich gut auf Basis einer additiven Integration beschreiben. Diese Resultate wurden in der vorliegenden Studie auf der Basis von theoretischen, neurowissenschaftlichen Ansätzen und psychophysiologischen Befunden diskutiert. Sie belegen die Flexibilität visueller Tiefenwahrnehmung und unterstreichen die Bedeutsamkeit individueller Merkmalseigenschaften sowie deren Interaktionen im Rahmen intrasensorischer Integrationsprozesse.