Computer Science and Artificial Intelligence Laboratory
Technical Report

MIT-CSAIL-TR-2010-046 September 21, 2010

Bayesian perceptual inference in linear Gaussian models

Peter W. Battaglia

Bayesian perceptual inference in linear Gaussian models

Peter W. Battaglia

September 21, 2010

BCS and CSAIL, MIT
Cambridge, MA 02139
pbatt@mit.edu

Abstract

The aim of this paper is to provide perceptual scientists with a quantitative framework for modeling a variety of common perceptual behaviors, and to unify various perceptual inference tasks by exposing their common computational underpinnings. This paper derives a model Bayesian observer for perceptual contexts with linear Gaussian generative processes. I demonstrate the relationship between four fundamental perceptual situations by expressing their corresponding posterior distributions as consequences of the model's predictions under their respective assumptions.

1 Introduction

Perception is the process of inferring scene properties that cannot be directly observed from the observable sensory evidence they generate. There is substantial regularity in the structure of scenes in the world, and sensations are generated by a consistent, yet noisy, process; modeling perception as Bayesian inference is attractive because it offers principled, normative behavioral predictions for perceptual tasks in this type of structured, uncertain world. Scientists compare these predictions with experimental measurements to reveal an organism's internal computational procedures responsible for its perceptually-guided behaviors.

However formulating and evaluating Bayesian perception models can be daunting due to their mathematical complexities and algorithmic subtleties. This report aims to increase perception scientists' access to Bayesian modeling by presenting a powerful class of models (linear Gaussian) that naturally encompass an important range of perceptual phenomena. These models are easy to construct, and can make exact, interpretable predictions.

Linear Gaussian models apply to perceptual situations in which the scene properties are apriori Gaussian-distributed, and generate sensory evidence by a linear process corrupted by Gaussian noise. They can also apply to log-Gaussian prior and noise distributions, with log-linear (multiplicative/divisive) sensory generative processes, and may be extended to more complicated generative processes for which Taylor series expansions provide good approximations. In practice, linear Gaussian models can be applied to many common sensation/perception situations, like spatial localization, temporal perception, size, lightness, and color constancy, and many others.

Section 2 presents the abstract linear Gaussian model and a deriviation of the posterior distribution over unobserved variables given observed variables. Section 3 tailors the framework to modeling four qualitatively-distinct, elementary perceptual situations [10]. Section 4 briefly outlines a decision-making framework compatible with the perceptual inference framework.

2 Linear Gaussian Model

This section presents the linear Gaussian model, and derives the posterior inference formulae. The derivation is based on general linear algebra rules, properties of Gaussians, and is examined in greater depth by [13].

2.1 Derivation of posterior for linear Gaussian model

Consider latent random vector, L, that generates observable data vector, D, as:

$$D = GL + \Omega$$

where G is a matrix that represents the deterministic component of the observation transform, Ω represents zero-mean ($\mathbf{0}$), additive observation noise, and $\mathcal{N}(X;Y,Z)$ is a normal distribution over X, with mean vector Y, and covariance matrix Z. Assume L and Ω have normal prior distributions,

$$\begin{aligned} \Pr(L) &= \mathcal{N}\left(L;\mu_L,\Sigma_L\right) \\ \Pr(\Omega) &= \mathcal{N}\left(\Omega;\mathbf{0},\Sigma_\Omega\right) \end{aligned}$$

By rules for linear transformations between normal random variables, the conditional and marginal likelihoods of the data are,

$$\begin{aligned} \Pr(D \mid L) &= \mathcal{N}\left(D;GL,\Sigma_\Omega\right) \\ \Pr(D) &= \mathcal{N}\left(D;G\mu_L,G\Sigma_L G^T + \Sigma_\Omega\right) \end{aligned}$$

Reformulating the conditional likelihood as an unnormalized distribution over L gives,

$$\Pr(D \mid L) = c \cdot \mathcal{N}\left(L;\left(G^T\Sigma_\Omega^{-1}G\right)^{-1}G^T\Sigma_\Omega^{-1}D,\left(G^T\Sigma_\Omega^{-1}G\right)^{-1}\right)$$

where c is a constant.

The joint distribution over L and D can be factored:

$$\Pr(L,D) = \Pr(D \mid L)\Pr(L)$$

and Bayes' theorem defines the posterior:

$$\begin{aligned} \Pr(L \mid D) &= \frac{\Pr(D \mid L)\Pr(L)}{\Pr(D)} \\ &= \frac{c \cdot k}{\Pr(D)}\mathcal{N}\left(L;\mu_{post},\Sigma_{post}\right) \end{aligned}$$

where k and $\Pr(D)$ are constants (μ_{post} and Σ_{post} are defined next).

Because $\Pr(L \mid D)$ and $\mathcal{N}(L; \mu_{post}, \Sigma_{post})$ are both densities, constant $\frac{c \cdot k}{\Pr(D)}$ must equal 1. So,

$$
\begin{aligned}
\Pr(L \mid D) &= \mathcal{N}(L; \mu_{post}, \Sigma_{post}) \\
\Sigma_{post} &= \left(G^T \Sigma_\Omega^{-1} G + \Sigma_L^{-1}\right)^{-1} \\
\mu_{post} &= \Sigma_{post} \left(G^T \Sigma_\Omega^{-1} D + \Sigma_L^{-1} \mu_L\right) \\
&= \left(G^T \Sigma_\Omega^{-1} G + \Sigma_L^{-1}\right)^{-1} G^T \Sigma_\Omega^{-1} D + \left(G^T \Sigma_\Omega^{-1} G + \Sigma_L^{-1}\right)^{-1} \Sigma_L^{-1} \mu_L
\end{aligned}
$$

If the posterior over only a subset of the elements of L is desired, because the posterior is normal the undesired latent elements can be easily marginalized out by deleting their corresponding rows (and columns) from μ_{post} and Σ_{post}.

3 Application to perceptual inference

This section considers several elementary perceptual situations, characterized by Kersten et al. (2004) [10] (Figure 4), by defining their generative processes under linear Gaussian assumptions, and their corresponding posterior inference rules. In each case, there are between 1 and 2 latent and data elements, but this can be extended arbitrarily by adding elements to the L and D vectors, and their respective parameter vectors/matrices. After each application, several references are provided in which the authors explicitly or implicitly use some form of the model.

3.1 Basic Bayes

Consider an observer who wishes to infer latent scene property $L = l$ from observed sensory data $D = d$, with $G = g$, $\Omega = \omega$, $\mu_L = \mu_l$, $\Sigma_L = \sigma_l^2$, and $\Sigma_\Omega = \sigma_\omega^2$.

The posterior over L given D is:

$$
\begin{aligned}
\Pr(L \mid D) &= \mathcal{N}(L; \mu_{post}, \Sigma_{post}) \\
\Sigma_{post} &= \left(\frac{g^2}{\sigma_\omega^2} + \frac{1}{\sigma_l^2}\right)^{-1} = \frac{\sigma_\omega^2 \sigma_l^2}{g^2 \sigma_l^2 + \sigma_\omega^2} \\
\mu_{post} &= \left(\frac{g^2}{\sigma_\omega^2} + \frac{1}{\sigma_l^2}\right)^{-1} \left(\frac{dg}{\sigma_\omega^2} + \frac{\mu_l}{\sigma_l^2}\right) = \frac{dg\sigma_l^2 + \mu_l \sigma_\omega^2}{g^2 \sigma_l^2 + \sigma_\omega^2}
\end{aligned}
$$

This model was used by [19] for modeling motion perception.

Using the common assumptions that $g = 1$ and $\sigma_l^2 \to \infty$ results in a simple perceptual inference rule,

$$\Pr(L \mid D) = \mathcal{N}(L; \mu_{post}, \Sigma_{post}) = \mathcal{N}\left(l; d, \sigma_\omega^2\right)$$

This special case has been used implicitly by too many authors to name, in a very broad number of perception studies.

3.2 Cue combination

Consider an observer who wishes to infer latent scene property $L = l$ from two pieces of observed sensory data $D = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$, with $G = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$, $\Omega = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}$, $\mu_L = \mu_l$, $\Sigma_L = \sigma_l^2$, and $\Sigma_\Omega = \begin{bmatrix} \sigma_{\omega_1}^2 & 0 \\ 0 & \sigma_{\omega_2}^2 \end{bmatrix}$.

The posterior over L given D is:

$$
\begin{aligned}
\Pr(L \mid D) &= \mathcal{N}(L; \mu_{post}, \Sigma_{post}) \\
\Sigma_{post} &= \left(\frac{g_1^2}{\sigma_{\omega_1}^2} + \frac{g_2^2}{\sigma_{\omega_2}^2} + \frac{1}{\sigma_l^2} \right)^{-1} = \frac{\sigma_{\omega_1}^2 \sigma_{\omega_2}^2 \sigma_l^2}{g_1^2 \sigma_{\omega_2}^2 \sigma_l^2 + g_2^2 \sigma_{\omega_1}^2 \sigma_l^2 + \sigma_{\omega_1}^2 \sigma_{\omega_2}^2} \\
\mu_{post} &= \left(\frac{g_1^2}{\sigma_{\omega_1}^2} + \frac{g_2^2}{\sigma_{\omega_2}^2} + \frac{1}{\sigma_l^2} \right)^{-1} \left(\frac{d_1 g_1^2}{\sigma_{\omega_1}^2} + \frac{d_2 g_2^2}{\sigma_{\omega_2}^2} + \frac{\mu_l}{\sigma_l^2} \right) \\
&= \frac{d_1 g_1^2 \sigma_{\omega_2}^2 \sigma_l^2}{g_1^2 \sigma_{\omega_2}^2 \sigma_l^2 + g_2^2 \sigma_{\omega_1}^2 \sigma_l^2 + \sigma_{\omega_1}^2 \sigma_{\omega_2}^2} + \frac{d_2 g_2^2 \sigma_{\omega_1}^2 \sigma_l^2}{g_1^2 \sigma_{\omega_2}^2 \sigma_l^2 + g_2^2 \sigma_{\omega_1}^2 \sigma_l^2 + \sigma_{\omega_1}^2 \sigma_{\omega_2}^2} + \frac{\mu_l \sigma_{\omega_1}^2 \sigma_{\omega_2}^2}{g_1^2 \sigma_{\omega_2}^2 \sigma_l^2 + g_2^2 \sigma_{\omega_1}^2 \sigma_l^2 + \sigma_{\omega_1}^2 \sigma_{\omega_2}^2}
\end{aligned}
$$

Using the common assumptions that $g_1 = g_2 = 1$ and $\sigma_l^2 \to \infty$ results in a simple perceptual inference rule,

$$
\Pr(L \mid D) = \mathcal{N}(L; \mu_{post}, \Sigma_{post}) = \mathcal{N}\left(l; \frac{d_1 \sigma_{\omega_2}^2 + d_2 \sigma_{\omega_1}^2}{\sigma_{\omega_1}^2 + \sigma_{\omega_2}^2}, \frac{\sigma_{\omega_1}^2 \sigma_{\omega_2}^2}{\sigma_{\omega_1}^2 + \sigma_{\omega_2}^2} \right)
$$

This model was used by [9, 20, 7, 4, 2, 12, 8] and many more for modeling human cue integration.

3.3 Discounting

Consider an observer who wishes to infer latent scene properties $L = \begin{bmatrix} l_1 \\ l_2 \end{bmatrix}$, from observed sensory data $D = d$, with $G = \begin{bmatrix} g_1 & g_2 \end{bmatrix}$, $\Omega = \omega$, $\mu_L = \begin{bmatrix} \mu_{l_1} \\ \mu_{l_2} \end{bmatrix}$, $\Sigma_L = \begin{bmatrix} \sigma_{l_1}^2 & 0 \\ 0 & \sigma_{l_2}^2 \end{bmatrix}$, and $\Sigma_\Omega = \sigma_\omega^2$.

The posterior over L given D is:

$$
\begin{aligned}
\Pr(L \mid D) &= \mathcal{N}(L; \mu_{post}, \Sigma_{post}) \\
\Sigma_{post} &= \begin{bmatrix} \frac{g_1^2}{\sigma_\omega^2} + \frac{1}{\sigma_{l_1}^2} & \frac{g_1 g_2}{\sigma_\omega^2} \\ \frac{g_1 g_2}{\sigma_\omega^2} & \frac{g_2^2}{\sigma_\omega^2} + \frac{1}{\sigma_{l_2}^2} \end{bmatrix}^{-1} \\
\mu_{post} &= \begin{bmatrix} \frac{g_1^2}{\sigma_\omega^2} + \frac{1}{\sigma_{l_1}^2} & \frac{g_1 g_2}{\sigma_\omega^2} \\ \frac{g_1 g_2}{\sigma_\omega^2} & \frac{g_2^2}{\sigma_\omega^2} + \frac{1}{\sigma_{l_2}^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{d g_1}{\sigma_\omega^2} + \frac{\mu_{l_1}}{\sigma_{l_1}^2} \\ \frac{d g_2}{\sigma_\omega^2} + \frac{\mu_{l_2}}{\sigma_{l_2}^2} \end{bmatrix}
\end{aligned}
$$

Using the common assumptions that $g_1 = 1$ and $\sigma_{l_1}^2 \to \infty$,

$$
\Pr(L \mid D) = \mathcal{N}(L; \mu_{post}, \Sigma_{post}) = \mathcal{N}\left(L; \begin{bmatrix} d - \mu_{l_2} g_2 \\ \mu_{l_2} \end{bmatrix}, \begin{bmatrix} g_2^2 \sigma_{l_2}^2 + \sigma_\omega^2 & -g_2 \sigma_{l_2}^2 \\ -g_2 \sigma_{l_2}^2 & \sigma_{l_2}^2 \end{bmatrix} \right)
$$

If only l_1 is relevant to the observer (i.e. l_2 is a "nuisance" variable), then,

$$\Pr(l_1 \mid D) = \mathcal{N}\left(l_1; \mu_{post}, \Sigma_{post}\right) = \mathcal{N}\left(l_1; d - \mu_{l_2} g_2, g_2^2 \sigma_{l_2}^2 + \sigma_\omega^2\right)$$

This model was implicitly used by [17, 16, 1], and many more, and can explain a variety of human perceptual biases.

3.4 Explaining away

Consider an observer who wishes to infer latent scene properties $L = \begin{bmatrix} l_1 \\ l_2 \end{bmatrix}$, from observed sensory data $D = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$, with $G = \begin{bmatrix} g_{1,1} & g_{1,2} \\ 0 & g_{2,2} \end{bmatrix}$, $\Omega = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}$, $\mu_L = \begin{bmatrix} \mu_{l_1} \\ \mu_{l_2} \end{bmatrix}$, $\Sigma_L = \begin{bmatrix} \sigma_{l_1}^2 & 0 \\ 0 & \sigma_{l_2}^2 \end{bmatrix}$, and $\Sigma_\Omega = \begin{bmatrix} \sigma_{\omega_1}^2 & 0 \\ 0 & \sigma_{\omega_2}^2 \end{bmatrix}$.

The posterior over L given D is:

$$\begin{aligned}
\Pr(L \mid D) &= \mathcal{N}\left(L; \mu_{post}, \Sigma_{post}\right) \\
\Sigma_{post} &= \begin{bmatrix} \frac{g_{1,1}^2}{\sigma_{\omega_1}^2} + \frac{1}{\sigma_{l_1}^2} & \frac{g_{1,1} g_{1,2}}{\sigma_{\omega_1}^2} \\ \frac{g_{1,1} g_{1,2}}{\sigma_{\omega_1}^2} & \frac{g_{1,2}^2}{\sigma_{\omega_1}^2} + \frac{g_{2,2}^2}{\sigma_{\omega_2}^2} + \frac{1}{\sigma_{l_2}^2} \end{bmatrix}^{-1} \\
\mu_{post} &= \begin{bmatrix} \frac{g_{1,1}^2}{\sigma_{\omega_1}^2} + \frac{1}{\sigma_{l_1}^2} & \frac{g_{1,1} g_{1,2}}{\sigma_{\omega_1}^2} \\ \frac{g_{1,1} g_{1,2}}{\sigma_{\omega_1}^2} & \frac{g_{1,2}^2}{\sigma_{\omega_1}^2} + \frac{g_{2,2}^2}{\sigma_{\omega_2}^2} + \frac{1}{\sigma_{l_2}^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{d_1 g_{1,1}}{\sigma_{\omega_1}^2} + \frac{\mu_{l_1}}{\sigma_{l_1}^2} \\ \frac{d_1 g_{1,2}}{\sigma_{\omega_1}^2} + \frac{d_2 g_{2,2}}{\sigma_{\omega_2}^2} + \frac{\mu_{l_2}}{\sigma_{l_2}^2} \end{bmatrix}
\end{aligned}$$

Using the common assumptions that $g_{1,1} = g_{2,2} = 1$, $\sigma_{l_1}^2 \to \infty$, and $\sigma_{l_2}^2 \to \infty$,

$$\Pr(L \mid D) = \mathcal{N}\left(L; \mu_{post}, \Sigma_{post}\right) = \mathcal{N}\left(L; \begin{bmatrix} d_1 - d_2 g_{1,2} \\ d_2 \end{bmatrix}, \begin{bmatrix} g_{1,2}^2 \sigma_{\omega_2}^2 + \sigma_{\omega_1}^2 & -g_{1,2}\sigma_{\omega_2}^2 \\ -g_{1,2}\sigma_{\omega_2}^2 & \sigma_{\omega_2}^2 \end{bmatrix}\right)$$

If only l_1 is relevant to the observer (i.e. l_2 is a "nuisance" variable), then,

$$\Pr(l_1 \mid D) = \mathcal{N}\left(l_1; \mu_{post}, \Sigma_{post}\right) = \mathcal{N}\left(l_1; d_1 - d_2 g_{1,2}, g_{1,2}^2 \sigma_{\omega_2}^2 + \sigma_{\omega_1}^2\right)$$

A slightly more general assumption is that $g_{1,1} = g_{2,2} = 1$, $\sigma_{l_1}^2 \to \infty$ ($\sigma_{l_2}^2$ remains finite),

$$\begin{aligned}
\Pr(L \mid D) &= \mathcal{N}\left(L; \mu_{post}, \Sigma_{post}\right) \\
\Sigma_{post} &= \frac{1}{\sigma_{l_2}^2 + \sigma_{\omega_2}^2} \begin{bmatrix} g_{1,2}^2 \sigma_{l_2}^2 \sigma_{\omega_2}^2 + \sigma_{l_2}^2 \sigma_{\omega_1}^2 + \sigma_{\omega_1}^2 \sigma_{\omega_2}^2 & -g_{1,2}\sigma_{l_2}^2 \sigma_{\omega_2}^2 \\ -g_{1,2}\sigma_{l_2}^2 \sigma_{\omega_2}^2 & \sigma_{l_2}^2 \sigma_{\omega_2}^2 \end{bmatrix} \\
\mu_{post} &= \begin{bmatrix} d_1 - \frac{d_2 g_{1,2} \sigma_{l_2}^2}{\sigma_{l_2}^2 + \sigma_{\omega_2}^2} - \frac{\mu_{l_2} g_{1,2} \sigma_{\omega_2}^2}{\sigma_{l_2}^2 + \sigma_{\omega_2}^2} \\ \frac{d_2 g_{1,2} \sigma_{l_2}^2}{\sigma_{l_2}^2 + \sigma_{\omega_2}^2} + \frac{\mu_{l_2} g_{1,2} \sigma_{\omega_2}^2}{\sigma_{l_2}^2 + \sigma_{\omega_2}^2} \end{bmatrix}
\end{aligned}$$

If only l_1 is relevant to the observer (i.e. l_2 is a "nuisance" variable), then,

$$\Pr(l_1 \mid D) = \mathcal{N}\left(l_1; \mu_{post}, \Sigma_{post}\right) = \mathcal{N}\left(l_1; d_1 - \frac{d_2 g_{1,2}\sigma_{l_2}^2 + \mu_{l_2} g_{1,2}\sigma_{\omega_2}^2}{\sigma_{l_2}^2 + \sigma_{\omega_2}^2}, \frac{g_{1,2}^2\sigma_{l_2}^2\sigma_{\omega_2}^2 + \sigma_{l_2}^2\sigma_{\omega_1}^2 + \sigma_{\omega_1}^2\sigma_{\omega_2}^2}{\sigma_{l_2}^2 + \sigma_{\omega_2}^2}\right)$$

This model was used by [5, 3, 6, 11], and can be applied to the broad class of perceptual constancy phenomena.

4 Perceptual decision-making

This section briefly outlines Bayesian decision theory, and how to apply the inference rules from Section 3 to make perceptual judgments. For a detailed treatment of Bayesian perceptual decision-making, see [14, 15].

4.1 Bayesian decision theory

Bayesian decision theory (BDT) prescribes optimal decision-making based on inferred posterior distributions over the state of the world. BDT defines a "risk" function, $\mathrm{R}(\alpha, D)$, that represents the expected reward (negative loss) for different combinations of data D and actions α. The risk function is an expectation over the observer's rewards, characterized by the reward (negative loss) function $\Lambda(\alpha, L)$, under the information inferred about the environment, characterized by the posterior $\Pr(L \mid D)$,

$$\mathrm{R}(\alpha, D) = \int_L \Lambda(\alpha, L)\Pr(L \mid D) dL$$

Optimal agents select actions with maximal expected reward by computing $\tilde{\alpha} = \arg\max_\alpha \mathrm{R}(\alpha, D)$. Sometimes random components, ν, are included to characterize the effects of motor noise, decision noise, etc., $\hat{\alpha} = \tilde{\alpha} + \nu$.

In normal psychophysical experiments, participants are typically asked to respond with an indication of the L state, so the space of α is either equivalent to the L space, or some straightforward function of it, $f(\cdot)$. In some cases they are assumed to place low, identical reward across all incorrect answers, and greater, identical reward across all correct answers:

$$\Lambda(\alpha, L) = \begin{cases} b & ; \alpha = f(\bar{L}) \\ c & ; \alpha \neq f(\bar{L}) \end{cases}$$

where $\bar{L}$ is the true latent world state and $b \gg c$.

For standard discrimination tasks, $f(\cdot)$ may be:

$$f(L) = \begin{cases} 0 & ; l_r \leq \theta \\ 1 & ; l_r > \theta \end{cases}$$

where l_r is an element of the latent state relevant to the task and θ is some threshold value (e.g. l_r represents absence/presence, 0/1, of an auditory tone and θ is 0).

For tasks that ask participants to produce a response, α, on some continuous axis, i.e. pointing, grasping, etc., $f(\cdot)$ may define a range, $[-\theta, \theta]$, that constitutes "success":

$$f(L) = \begin{cases} 1 & ; -\theta \leq l_r \leq \theta \\ 0 & ; \text{otherwise} \end{cases}$$

In work from the cognitive psychology domain and more recent perceptual studies [21], it has been suggested that humans sample from their posterior distributions, rather than computing deterministic functions of the distribution, to produce responses. Also, in some cases the perceptual inference process may appear to depend on the task [18].

5 Conclusion

The benefits of a Bayesian framework for modeling perceptually-guided behaviors is that it explains these behaviors as resulting from a principled inferential process, based on sensory input and the observer's perceptual system's internal knowledge and beliefs, which is used rationally to acquire reward and avoid penalty. The linear Gaussian model and its example applications provide an accessible and useful model that can be used to model a broad set of perceptual behaviors. It may help explain various cue integration, bias, constancy phenomena.

The problem of using this framework to design an experiment and analyze the resultant data is beyond the scope of this article, but is an important element for leveraging this framework on typical perceptual science problems. In brief, by treating the experimental stimuli as L (Section 2.1), and the recorded participant responses as $\hat{\alpha}$ (Section 4.1), the model parameters, $(G, \mu_L, \Sigma_L, \Sigma_\Omega \nu)$, which encode the experimenter's hypotheses and assumptions about the observer's perceptual reasoning, can be estimated or inferred using standard statistical methods.

Acknowledgements

Paul Schrater, Frank Jaekel, and Dan Kersten provided helpful direction and criticism. Funding: NIH Grant (NRSA) Number F32EY019228-02, and a UMN Doctoral Dissertation Fellowship.

References

[1] W.J. Adams, E.W. Graf, and M.O. Ernst. Experience can change the'light-from-above'prior. *Nature Neuroscience*, 7(10):1057–1058, 2004.

[2] D. Alais and D. Burr. The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14(3):257–262, 2004.

[3] P.W. Battaglia, M. Di Luca, M.O. Ernst, P.R. Schrater, T. Machulla, and D. Kersten. Within-and Cross-Modal Distance Information Disambiguate Visual Size-Change Perception. 6(3):e1000697, 2010.

[4] P.W. Battaglia, R.A. Jacobs, and R.N. Aslin. Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, 20(7):1391–1397, 2003.

[5] P.W. Battaglia, P.R. Schrater, and D.J. Kersten. Auxiliary object knowledge influences visually-guided interception behavior. In *Proceedings of the 2nd symposium on Applied perception in graphics and visualization*, volume 95, pages 145–152. ACM, 2005.

[6] MG Bloj, D. Kersten, and AC Hurlbert. Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*, 402(6764):877–879, 1999.

[7] M.O. Ernst and M.S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002.

[8] J.M. Hillis, S.J. Watt, M.S. Landy, and M.S. Banks. Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12), 2004.

[9] R.A. Jacobs. Optimal integration of texture and motion cues to depth. *Vision Research*, 39(21):3621–3629, 1999.

[10] D. Kersten, P. Mamassian, and A. Yuille. Object perception as Bayesian inference. *Annu. Rev. Psychol*, 55:271–304, 2004.

[11] D.C. Knill and D. Kersten. Apparent surface curvature affects lightness perception. *Nature*, 351(6323):228–230, 1991.

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

[13] DV Lindley and AFM Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–41, 1972.

[14] L.T. Maloney. Statistical decision theory and biological vision. *Perception and the physical world: Psychological and philosophical issues in perception*, pages 145–189, 2002.

[15] L.T. Maloney and P. Mamassian. Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual neuroscience*, 26(01):147–155, 2009.

[16] P. Mamassian and R. Goutcher. Prior knowledge on the illumination position. *Cognition*, 81(1):B1–B9, 2001.

[17] P. Mamassian and M.S. Landy. Interaction of visual prior constraints. *Vision Research*, 41(20):2653–2668, 2001.

[18] P.R. Schrater and D. Kersten. How optimal depth cue integration depends on the task. *International Journal of Computer Vision*, 40(1):71–89, 2000.

[19] A.A. Stocker and E.P. Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4):578–585, 2006.

[20] R.J. van Beers, A.C. Sittig, and J.J. Gon. Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, 81(3):1355, 1999.

[21] E. Vul, N.D. Goodman, T.L. Griffiths, and J.B. Tenenbaum. One and done? Optimal decisions from very few samples. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society, Amseterdam, the Netherlands*, 2009.