Uncertainty, generalization, and neural representation of relevant variables for decision making

Hugo L. Fernandes

Dissertation presented to obtain the

Ph.D degree in Biology | Computational Biology

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Research work coordinated by:

FUNDAÇÃO CALOUSTE GULBENKIAN
Instituto Gulbenkian de Ciência

Oeiras,

December, 2013

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

Apoio Financeiro/Financial Support

Apoio financeiro da FCT e do FSE no âmbito do Quadro Comunitário de Apoio, bolsa de doutoramento n° SFRH/BD/33525/2008.

Financial support for this thesis was provided by FCT and FSE through the Quadro Comunitário de Apoio, doctoral fellowship n° SFRH/BD/33525/2008.

List of Publications Included in the Thesis

This thesis is based on the following publications:

Chapter 2:
Fernandes HL, Stevenson IH, Körding KP (2012)
Generalization of stochastic visuomotor rotations. PLoS ONE

Chapter 3:
Fernandes HL, Stevenson IH, Körding KP (under review)
The generalization of prior uncertainty during reaching.

Chapter 4:
Acuna DE*, Berniker M*, Fernandes HL*, Körding KP (in preparation)
Prior beliefs and decision-making under uncertainty
*contributed equally

Chapter 5:
Fernandes HL, Phillips AN, Stevenson IH, Segraves MA, Körding KP (2013)
Saliency and saccade encoding in the frontal eye field during natural scene search. Cerebral Cortex

List of Publications not Included in the Thesis

Cherian A, **Fernandes HL**, Miller LE (2013)
Primary motor cortical discharge during force field adaptation reflects muscle-like dynamics. Journal of Neurophysiology

Vilares I, Howard JD, **Fernandes HL**, Gottfried J and Körding K (2012)
Prior and likelihood uncertainty are differentially represented in the human brain. Current Biology

Avraham G, Nisky I, **Fernandes HL**, Acuna DE, Körding KP, Feldman AG, Loeb GE, Karniel A (2011)
Towards Perceiving Robots as Humans – Three handshake models face the Turing-like Handshake Test. IEEE Transactions on Haptics

Fernandes HL*, Albert MV*, Körding KP (* contributed equally) (2011)
Measuring generalization of visuomotor perturbations in wrist movements using mobile phones. PLoS ONE

Fernandes HL, Körding KP (2010)
In praise of “false” models and rich data. Journal of Motor Behavior

Stevenson IH, **Fernandes HL**, Vilares I, Wei K, and Körding KP (2009)
Bayesian integration and non-linear feedback control in a full-body motor task. PLoS Computational Biology

Summary

Decision making is a ubiquitous theme in computational neuroscience. Here we present novel data and modeling approaches that relate to how the brain makes decisions. In the first part of this dissertation we use psychophysics to look into how the brain itself does statistics; in the second part we use statistical analysis tools to investigate the neural representation of relevant variables for decision making.

Innumerous studies have suggested that people take previously accumulated information (prior distribution) as well and new information (the likelihood) into account when making a decision. Here we start by asking where do prior distributions come from; since we are rarely in the exact same situation twice, how is the prior, used in a particular decision, generalized from previous similar experiences? Using a movement experiment we found differences between the generalization of the mean and variance of the prior distribution.

We continue by asking how the brain makes a decision when choosing between two alternatives in a two-alternative-forced-choice (2AFC) task. The 2AFC paradigm is often assumed to measure sensory (likelihood) uncertainty independently of prior uncertainty. Here we test this assumption by looking into the algorithms the brain might use when choosing between the two alternatives in a 2AFC task. Specifically, after combining the prior and likelihood into a posterior distribution, is the decision based on the maximum of the posterior (the MAP hypothesis) or do humans sample from the posterior distribution (the sampling/matching hypothesis)? We show that in investigating this question we simultaneously test whether the 2AFC paradigm can be used to measure sensory uncertainty independently of prior uncertainty. Our experimental results favor the MAP hypothesis and hence the validity of the assumption.

Finally, we use probabilistic models to investigate whether neurons in the frontal eye field (FEF) represent bottom-up saliency when monkeys are searching a natural scene. Understanding what the brain represents/does ultimately involves understanding how it represents the kind of stimuli it has evolved to represent. Here we use natural scenes and an objective definition of bottom-up saliency that has been shown to predict saccade choices of both humans and monkeys during free-viewing of natural scenes. We found that although saliency appears to be used in deciding where to look next and predicts neural activity of FEF neurons, its predictive power is explained away if we take into account other saccade related covariates.

This thesis provides important insights into several aspects of decision making. At a higher level it provides data for constraining models of generalization of uncertainty; it tests theories that relate to which kind of decision-making algorithms the brain implements; and finally it looks into neural representation of natural stimuli that are relevant for decision making.

Resumo

A tomada de decisão é um tema omnipresente em neurociência computacional. Aqui apresentamos novas abordagens e resultados relacionados com a forma como o cérebro toma decisões .

Começamos por fazer perguntas relacionadas com o modo como o próprio cérebro faz estatística. Especificamente, perguntamos de onde vêm as probabilidades a priori - as distribuições de probabilidade sobre o que esperamos que aconteça. Uma vez que raramente nos encontramos na mesma exacta situação duas vezes, como é que estas distribuições, utilizadas numa decisão particular, são generalizadas a partir de experiências semelhantes anteriores? Usando uma experiência de movimento encontramos diferenças de abrangência e simetria entre a generalização da média e variância da distribuição de probabilidade a priori.

De seguida perguntamos como é que o cérebro toma uma decisão a partir da distribuição de probabilidade a posteriori quando tem de escolher entre duas alternativas no paradigma two alternative forced choice (2AFC). Usa o máximo da distribuição ou amostra da distribuição? Mostramos que ao investigar esta questão estamos simultaneamente a testar a hipótese comumente usada, mas não testada sobre este paradigma: que pode ser usado para medir a incerteza sensorial independentemente de incerteza das expectativas, i.e. independentemente da incerteza na distribuição a priori.

Finalmente, usamos modelos probabilísticos para investigar as representações de variáveis relevantes para a tomada de decisão. Especificamente, perguntamos se o frontal-eye-field (FEF) de macacos representa a saliência de imagens quando os macacos estão à procura de um objecto em imagens naturais. Mostramos que, embora a saliência pareca ser usada para decidir para onde olhar e prevê a actividade de

neurónios no FEF, o seu poder preditivo desaparece (ou é explicado) se tomarmos em consideração variáveis relacionadas com o movimento dos olhos.

Contents

1. General Introduction

Understanding decision making in various contexts is fundamental to understanding human behavior. This thesis presents several studies that examine decision making from many different points of view using a variety of research tools.

In Chapters 2-4, we use human psychophysics, i.e., behavioral experiments designed for the quantitative study of the perceptual system. We use some of these experiments to characterize the generalization of prior expectations/subjective beliefs and to investigate which algorithms the nervous system uses for making a decision. In Chapter 5, we analyze neural recordings to understand the representation of relevant neural variables for eye-movement decision making. The experiments presented here cover a wide range of decisions including motor decisions about where to reach (Chapters 2 and 3), sensory discrimination decisions when confronted with two choices (Chapter 4), and attention-related oculomotor decisions about where to look next (Chapter 5).

1.1 Bayes theorem, priors, likelihood and posterior

Innumerous studies have suggested that, when making a decision, humans take previous accumulated information (their expectations, or the prior distribution) as well as new information (the likelihood) into account (for a review see Vilares and Kording, 2011). Both of these pieces of information have an associated mean and variance (denoted by uncertainty). This perspective/modeling approach/description is called Bayesian and owes its name to the Bayes theorem:

$$p(r|s)=\frac{p(s|r)p(r)}{p(s)}$$

When we try to infer what is out there in the world, we are interested is in having an accurate measure of what the current *reality* is given the new information arriving to our *senses* ($p(r|s)$, *the posterior*). Bayes theorem tells us how we can obtain it, i.e., how to combine the new piece of sensory information ($p(s|r)$, the *likelihood*) with our prior expectations over reality (the prior, $p(r)$).

1.2 Generalization, where priors come from

Where do priors come from? As we are never in the exact same situation twice, it is useful to generalize about subjective beliefs acquired in one situation to be applicable to different but similar situations (Shepard, 1987). In Chapters 2 and 3 we address the question of how priors generalize and, in particular, how prior variance/uncertainty generalizes. Several neural representation theories have been proposed on how the brain might represent/approximate the prior and the likelihood (Deneve, 2008; Fiser et al., 2010a; Hinton and Sejnowski, 1983a; Hoyer and Hyvärinen, 2003; Ma et al., 2006; Ma, 2010; Sahani and Dayan, 2003; Soltani and Wang, 2009; Wu et al., 2003; Zemel et al., 1998). These theories propose diverse, but not always mutually exclusive, ways in which uncertainty could be represented by populations of neurons. Some propose that it is represented in the width of the tuning curves, others in the amplitude of the tuning curves, in the firing rate, in the strength of the synapses, in the timing of the firing, etc. To our knowledge however, none of theories about the neural representation of uncertainty has been extended to incorporate generalization, nor is there data on how the variance of the prior distribution generalizes; generalization studies typically neglect variability.

Our aim was thus to characterize how the prior generalizes, and specifically to understand how this generalization depends on uncertainty. Namely, we wanted to understand: 1. how the generalization of the mean of the prior is affected by different degrees of uncertainty; and 2. how the variance/uncertainty of the prior itself generalizes. For that, we used a previously established generalization paradigm (Krakauer et al., 2000), which consists on a visuomotor rotation during a center-out reaching task, and extended it to include uncertainty/variability (Körding and Wolpert, 2004).

We found that manipulating the uncertainty level (the variance) of the prior does not affect how the mean of the prior generalizes (Chapter 2). We find differences in breadth between the generalization of mean and variance (uncertainty) and an unexpected asymmetry in the generalization of uncertainty (Chapter 3). Using a gradient-descent model we find that this asymmetry is consistent with the use of different similarity reference frames between the generalization of mean and variance/uncertainty. The results from Chapter 2 and 3 characterize differences and similarities between the generalization patterns of mean and uncertainty of prior expectations, and constrain future extensions of theories of prior representation to include effects of learning and generalization.

1.3 How are decisions made?

Although we know that Bayesian decisions involve a combination of prior and likelihood information, there exist several mathematical strategies by which an inference/decision could be computed. In Bayesian decision-theory, after arriving to the posterior distribution, the ultimate decision still depends on the cost/reward of each of the possible choices. If everything (prior and likelihood) is assumed to be Gaussian, and under reasonable

choices of cost function (e.g. the mean squared error), the ideal choice is to weight the mean of prior and the mean of the likelihood by their relative precisions (the reciprocal of the variances). This choice corresponds to choosing the maximum of the posterior distribution (MAP). Do Humans use this strategy when deciding between two choices?

In Chapter 4 we investigate which strategy humans use when deciding between two possible choices. Specifically we ask whether the final decision is based on the maximum of the posterior distribution (the MAP hypothesis), or do humans instead sample from the posterior distribution (the sampling and matching hypothesis) (Vul et al., 2009; Vulkan, 2000; Wozny et al., 2010). To investigate this we use the *two-alternative forced choice paradigm* (2AFC), a discrimination task that is one of the most used paradigms in psychophysics.

1.4 Two-alternative forced choice paradigm and the just-noticeable difference

In a 2AFC task, subjects are presented with two alternatives and forced to choose between them. For example, subjects may be asked to decide which of two tones has a higher pitch. By controlling the discrepancy between these tones (the cues), experimenters can obtain a psychometric curve: the probability of a subject's response given the discrepancy between cues. This curve is often used to quantify the *just-noticeable difference* (JND), which is related to how different the two cues must be before subjects can tell them apart.

The JND is often assumed to measure sensory uncertainty, i.e., uncertainty/variance of the likelihood, independently of the variance of the prior. The MAP decision-making hypothesis described above is the implicit and untested assumption in studies that use JND for measuring likelihood uncertainty -- or at least consistent with that objective. However, as we show in Chapter 4, if the sampling/matching hypotheses are true then the JND is in fact proportional to perceptual uncertainty (i.e. proportional to the variance of the posterior distribution). Importantly this would mean that the JND is affected by changes in prior uncertainty and hence that it should be used with caution. The prevalence of either of these hypotheses has thus broad implications for the interpretability of the 2AFC paradigm.

In Chapter 4 we present a task that allows manipulation of subjects' prior uncertainty while simultaneously measuring subjects' JND. Our results suggest that prior uncertainty does not affect the subjects' JND. Hence our results support the MAP hypothesis and the use of JND to measure sensory uncertainty. Importantly we show how the 2AFC task can be used to test these decision-making theories.

1.5 Deciding where to look next

A big part of our sensory information comes through our retina. When we are scanning a visual scene, we are constantly moving our eyes from one place to the next. In fact, deciding where to look next might be one of our most frequent decisions. How do we accomplish it? In order to understand how we chose where to look next, the computational modeling of eye-fixation choices has found that both bottom-up/task independent image features such as bottom-up saliency (Itti and Koch, 2001), as well as top-down features, such as target similarity (Einhäuser et al., 2008) can predict eye movements to some degree.

In the second part of the thesis (Chapter 5) we look into neural representation of one of these important variables for deciding where to look next; bottom-up saliency (Itti and Koch, 2001). We search for neural representation of bottom-up saliency in the frontal-eye-field (FEF) while monkeys are searching natural scenes for an embedded target (Phillips and Segraves, 2010). The FEF is a brain region that is thought to be involved in the production of saccades, while at the same time responding to salient visual stimuli. However, while some experiments using artificial stimuli suggest that saliency is represented in the FEF, understanding what the brain represents/does ultimately involves understanding how it represents the kind of stimuli it has evolved to represent (Kayser et al., 2003; MacEvoy et al., 2008; Theunissen et al., 2000). We use natural scenes and an objective definition of bottom-up saliency that has been shown to predict both human and monkey's saccade choices. We find that basic analyses suggest that FEF represents both saccade direction and saliency. However, by using Generalized Linear Models (Pillow et al., 2008; Saleh et al., 2010; Truccolo et al., 2005), specifically linear–nonlinear-Poisson cascade models, we show that saccade covariates explain away (Pearl, 1988) bottom-up saliency. Hence, even though saliency appears to be used when deciding where to look next, it does not seem that FEF neurons actively represent it during natural scene search.

1.6 Marr's levels in this dissertation

To understand and contextualize the contributions of this dissertation in a unified sense, it is useful to group them under Marr's levels of analysis. David Marr (Marr, 1982) introduced a taxonomy of three different levels of description/analysis. According to Marr, it is possible to divide models into those that deal with the objective of computation (Level 1), the algorithm used (Level 2), and the implementation (Level 3). The Level 1 approaches

ask which computational problem the nervous system is trying to solve. Normative approaches, which ask which kind of computation the brain should be solving, are included in the category. A typical example is the question of if the nervous system combines cues from different modalities taking into account their uncertainty to obtain a minimum variance estimate (Kording, 2007). Level 2 models deal with which algorithm or which strategies does the nervous system use to solve the computational objective. Finally level 3 deals with the precise physical implementation of the level 2 algorithms. The implementation can itself be described at many levels: molecular, synapses, spikes, etc. A description at level 1 can have several descriptions at level 2, and an algorithm at level 2 can have several possible implementations at level 3. However a particular implementation should originate one algorithm and a particular algorithm is typically solving one particular computational objective. Research done at a particular level thus constrains the possible descriptions not only at that level but also at the other levels and one could argue that the richer approaches are usually the ones able to connect different levels.

The research presented in this dissertation touches several Marr levels. While Chapter 2 and 3 are mostly experimental, in these chapters we examine the learning and generalization of the prior distribution assuming a normative model of decision making: Bayesian decision theory (Marr level 1). In Chapter 4 we test decision-making algorithms (Marr's level 2) while also using the normative approach of Bayesian decision theory. Finally in Chapter 5 we use a computational definition of natural scene saliency that is generally used to predict eye movements. Hence, we test a specific algorithm (Marr's level 2) for how the brain decides where to look next. While it is difficult to say exactly what mechanism/implementation means in Marr's level 3 — it constitutes many levels of explanation — Chapter 5 investigates whether and how features of a computational algorithm (Marr's

level 2) for deciding where to look next is implemented in a specific population of neurons (Marr's level 3).

2. Generalization of Stochastic Visuomotor Rotations

Hugo L. Fernandes, Ian H. Stevenson and Konrad P. Kording

Citation. Fernandes HL, Stevenson IH, Kording KP (2012) Generalization of Stochastic Visuomotor Rotations. PLoS ONE 7(8): e43016. doi:10.1371/journal.pone.0043016

Author Contributions. Conceived and designed the experiments: HLF IHS KPK. Performed the experiments: HLF. Analyzed the data: HLF IHS KPK. Implemented the analysis: HLF with the contribution of IHS and KPK, Contributed reagents/ materials/ analysis tools: HLF IHS KPK. Wrote the paper: HLF IHS KPK

2.1 Summary

Generalization studies examine the influence of perturbations imposed on one movement onto other movements. The strength of generalization is traditionally interpreted as a reflection of the similarity of the underlying neural representations. Uncertainty fundamentally affects both sensory integration and learning and is at the heart of many theories of neural representation. However, little is known about how uncertainty, resulting from variability in the environment, affects generalization curves. Here we extend standard movement generalization experiments to ask how uncertainty affects the generalization of visuomotor rotations. We find that although uncertainty affects how fast subjects learn, the perturbation generalizes independently of uncertainty.

2.2 Introduction

A central goal of systems neuroscience in general and motor control research in particular is to understand how sensorimotor behaviors, such as reaching, are represented and learned. One factor that regularly influences movement planning and execution is uncertainty. For example, when we grasp objects our hands move very differently depending on our level of uncertainty; if we are uncertain about an object's position, we open our hands wider, move more slowly and approach the object with our hands aligned with the direction of highest uncertainty (Christopoulos and Schrater, 2009). This example highlights the fact that variability in the external world affects behavior and suggests that uncertainty must be represented in the nervous system.

Many studies in the field of motor control have used generalization experiments to examine the neural representation of movement, asking how learning a perturbation in one task affects behavior on novel tasks (Donchin et al., 2003; Ghahramani et al., 1996b; Goodbody and Wolpert, 1998a; Hwang et al., 2006; Krakauer et al., 2000a; Mattar and Ostry, 2007; Paz et al., 2003; Pearson et al., 2010; Shadmehr, 2004; Shadmehr and Moussavi, 2000; Shadmehr and Mussa-Ivaldi, 1994; Thoroughman and Shadmehr, 2000; Thoroughman and Taylor, 2005). By studying which aspects of the behavior are transferred between tasks and which tasks a behavior transfers to, these experiments have investigated how we represent and modify movement and task variables. Generalization is sensitive to many factors including the coordinate system, nature, and complexity of the perturbation (Hwang et al., 2006; Krakauer et al., 2000a; Shadmehr and Mussa-Ivaldi, 1994; Thoroughman and Taylor, 2005), movement variables such as speed (Goodbody and Wolpert, 1998) and posture (Shadmehr and Moussavi, 2000), as well as the extent and type of training and feedback (Pearson et al., 2010; Taylor et al., 2012). However, one factor that has not

yet been studied in the context of generalization experiments is uncertainty. Many studies have explored how uncertainty affects behavior (Christopoulos and Schrater, 2009; Körding and Wolpert, 2004a; Saijo and Gomi, 2012; Tassinari et al., 2006), but how uncertainty influences generalization has received little attention.

From a normative viewpoint, subjects should generalize what they have learned about a perturbation in one situation to a novel situation only if they expect the perturbation to occur in the novel situation. Behavior in novel situations reveals what subjects expected to occur, and these expectations may be affected by several factors including task similarity or familiarity with the type of perturbation. It has been difficult to formalize this normative approach to generalization, since natural movement statistics and natural perturbation statistics are difficult to collect. However, any normative description of generalization must take uncertainty into account, since variability in the external world can have strong effects on behavior; task uncertainty (Körding and Wolpert, 2004), sensory uncertainty (Wei and Körding, 2010) and motor noise (Harris and Wolpert, 1998; van Beers, 2009), have all been shown to affect individual movements and learning, and may affect the similarity between movements as well as the resulting generalization.

A common interpretation of generalization from one task to another is that stronger generalization indicates a larger overlap in the neural representations of the two tasks. For instance, Krakauer et al. (Krakauer et al., 2000) measured generalization of planar, center-out reaching movements with rotation and gain perturbations. Training with a rotational perturbation in one direction produced strong generalization to nearby angular targets, but did not affect movements to novel targets with large angular separations from the training direction (>45°). On the other hand, visuomotor gain perturbations tended to generalize globally, to all reach directions. This finding suggests that the internal neural representation that

changed in response to these perturbations is activated during movements to similar angular directions, and that there may be a polar representation of planar reaches, where reach angle and extent are independent. Here we extend a visuomotor rotations experiment of Krakauer et al. (Krakauer et al., 2000) by introducing variability in the perturbations.

It is not clear, a priori, if and how uncertainty might influence generalization. One hypothesis, from a normative perspective, might be that task variability will make subjects more conservative and generalization narrower. High variability may indicate to subjects that it is less likely that the perturbation will be present for novel targets. A second hypothesis is that higher uncertainty will result in broader neural representations and that these could be reflected in wider generalization patterns. Several theories of the neural representation of uncertainty explicitly predict that uncertainty changes neural tuning. In particular, these models predict that tuning of individual neurons becomes wider with higher uncertainty (Girshick et al., 2011; Zemel et al., 1998), and there is some experimental data suggesting that this may be the case (Barlow et al., 1957; Cisek and Kalaska, 2005) (see Discussion). If generalization patterns trivially reflect overlapping neural tuning and if neural tuning becomes wider with increasing uncertainty then we might expect generalization to become broader with increasing uncertainty. However, it is difficult to match behavioral results to precise neural mechanisms; generalization between two movements can typically only be interpreted in terms of the degree of behavioral similarity between the movements or in terms of an abstract similarity between the neural representations of the two movements (Poggio, 1990; Poggio and Bizzi, 2004; Pouget and Snyder, 2000; Thoroughman and Shadmehr, 2000).

Here, with the goal of examining how uncertainty influences generalization patterns, we designed an experiment to manipulate the mean and the variance of noisy visuomotor rotations relative to the central starting position while subjects performed center-out reaches. We examined how

subjects adapt to distributions of perturbations applied during movement in one direction (training direction). On each trial a rotation sampled from a Gaussian distribution with fixed mean and variance was applied to a hidden cursor controlled by the subjects' index finger. After the subjects adapt, we measure how the learned mean generalizes to movements into other directions. The mean perturbation remained the same under the different noise conditions. If this mean is the only factor driving generalization movements, then we would not expect to see any difference between generalization curves. On the other hand, since uncertainty has been shown to affect many different types of movement, it is important to test whether or not generalization changes under noisy perturbations. We found that the mean of the perturbation generalizes with a width of about 30 degrees, in line with previous studies (Fernandes et al., 2011; Krakauer et al., 2000a; Paz et al., 2005). We found that the variance of the perturbation changes the speed and extent of learning, but, importantly, generalization is unaffected.

2.3 Results

Here we ask how a perturbation that varies randomly across trials is learned for one direction and how adaptation to this perturbation affects movements into other directions. We thus extend movement generalization studies by analyzing how uncertainty, induced by variability or noise in the perturbation, affects generalization patterns. Subjects controlled the position of a hidden cursor with their right index finger by making planar reaches in a projector-mirror system that blocked the view of the hand (**Figure 2.1A**). They made center-out reaches from the workspace center to one of eight targets while a visuomotor rotation, relative to the workspace center position, was applied to the hidden cursor position. The visuomotor rotation was drawn randomly each trial from a Gaussian distribution with fixed mean and variance (**Figure**

2.1B). During *learning* subjects were incentivized to make reaches to one of the targets and received endpoint feedback about the cursor position that allowed them to adapt to the perturbations. During *testing* subjects made reaches to the other targets, without endpoint feedback, allowing us to examine the generalization patterns (**Figure 2.1C**). We then measured how learning about the rotations under different variance conditions generalized.

Subjects (n=16) were confronted with a rotational perturbation that caused the cursor to deviate from the true hand position as subjects moved away from the center of the workspace. We presented three blocks of training with the same absolute mean perturbation (±30 degrees) but different variability (standard deviations, σ_p : 0°, 4° or 12°). Since the sign of the mean of the perturbation was randomly chosen for each subject and condition, in order to compare across subjects we transformed the measure of generalization so that positive hand position angles always refer to hand position angles that counteract the average perturbation – we call this measure the *absolute angle of final hand position*. In agreement with previous studies (Berniker et al., 2010b; Burge et al., 2008), we found that subjects rapidly adapt to the mean rotation, and, while they initially make large errors, subjects learn to counter-act the perturbation so that errors become small over the course of a few trials (**Figure 2.2**). We found that learning is fastest ($p<0.03$, bootstrap) and most complete ($p<0.001$, bootstrap) for the condition with zero variance (see Methods for details). As the uncertainty of the perturbation increased learning was both slower and less complete.

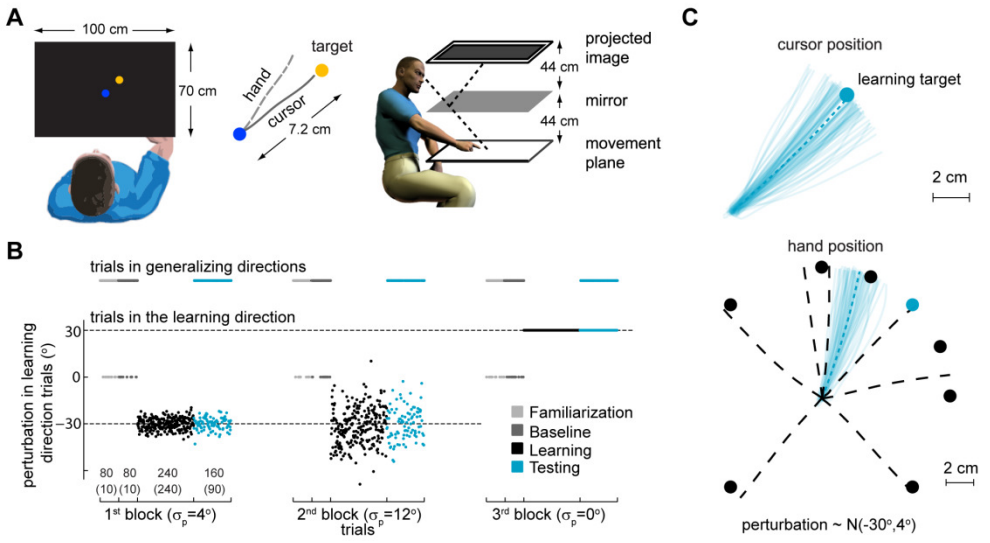


Figure 2.1. Experimental setup, protocol and typical trajectory data. A) Experimental setup. Subjects control the position of a hidden cursor on the screen with their right index finger. A projector-mirror system allows the image on-screen to be perceived as being in the movement plane. Subjects were incentivized to reach to a target (yellow) starting from a central target position (blue). The experiment assesses generalization of the learned mean under different uncertainty conditions. B) Perturbations and block design for an individual subject. Sequence of trials in the learning direction and generalizing directions and perturbations applied to trials in the learning direction for an individual subject. σ_p denotes the standard deviation of the distribution of perturbations. Each block is composed of 4 sub-blocks: *familiarization, baseline, learning* and *testing*. Numbers in the 1st block horizontal axis correspond to the total number of trials during each sub-block (no brackets) and the number of trials towards the learning direction during each sub-block (between brackets). C) Typical hand and cursor position during a *testing* sub-block. Thin colored lines are movements towards the learning target (colored circles). Dashed thick lines are average hand position for reaches in each direction. Black circles are targets in generalizing directions.

Once subjects learn the perturbation in one direction we assess how this learned perturbation generalizes. Using the average final hand position during movements to the testing directions as a measure of generalization, we found that the generalization patterns are local in all three variance conditions (**Figure 2.3A-C**). This is in line with Krakauer et al. (Krakauer et al., 2000) whose main condition was essentially identical to our $\sigma_p = 0^o$ condition. Given that different subjects have different baseline biases and the amount of learning changes depending on subject and condition, we subtracted the baseline biases and normalized the generalization by the amount of learning in the learning direction – we call this measure the *percent adaptation* relative to the learning direction (**Figure 2.3C**, see Methods). Despite the fact that uncertainty influenced the rate and amount of adaptation, we did not find a difference between the generalization curves in the three conditions in the absolute angle of final hand position (**Figure 2.3B**) ($F_{2,210}$=1.06, p=0.36, two-way repeated measures ANOVA) or in the percent adaptation relative to the learning direction (**Figure 2.3C**) ($F_{2,210}$=0.11, p=0.89, two-way repeated measures ANOVA). We also did not find a significant interaction between uncertainty levels and target angle either in the absolute angle of final hand position ($F_{14,210}$=1.31, p=0.20, two-way repeated measures ANOVA) or in the percent adaptation relative to the learning direction ($F_{14,210}$=0.63, p=0.84, two-way repeated measures ANOVA). These results suggest that the generalization pattern is independent of the uncertainty about the perturbation.

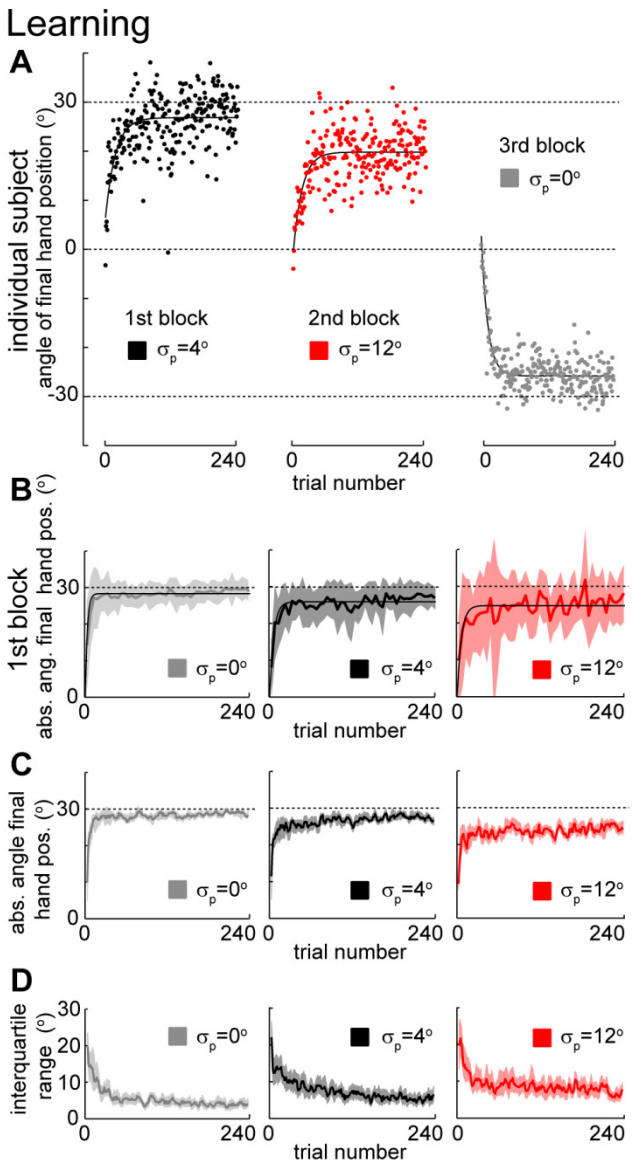


Figure 2.2. Learning of mean under different variance conditions. A) Learning the mean of a perturbation during the first perturbation block for a typical subject. Solid lines denote exponential fits. B) Learning the mean of a perturbation during the first perturbation block across subjects (n=8, n=4, and n=4 for the standard deviations, σ_p of 0°, 4° and 12°, respectively). Thick lines are average (±SD) across subjects considering bins of 5 trials. Thin lines are exponential fits. Grey dashed lines indicate the absolute average of the imposed perturbation (30°). C) Learning the

mean of a perturbation considering all blocks for each variance condition. Thick lines denote medians across subjects and trials in a trial window of 5 trials. Shaded area is 95% confidence region (bootstrap). D) Variability of angle of final hand position. Thick lines denote the interquartile range of the angle of final hand position across subjects and a trial window of 5 trials. Shaded area is 95% confidence region (bootstrap).

With the exception of the transformed sign of the angle of final hand position (for the measures absolute angle and percent adaptation), we have thus far ignored the sign of the perturbation ($+30^o$ or -30^o) in our analysis. We can take the sign of the perturbation it into account by reflecting the target directions (x-axis in **Figure 2.3A-C**) of the generalization data relative to the learning target direction for those blocks that had a -30^o as mean of the distribution of perturbations. Given that all sixteen subjects are right-handed, by ignoring the sign of the mean of the distribution of perturbations while combining the data from the different subjects we expect to detect biomechanical biases that could eventually scale with the level of variability but independently of the sign of the perturbation (**Figure 2.3B-C**). On the other hand if we take into account the sign of the perturbation we test for influences of angular direction of the mean of the perturbations on generalization and how these might eventually scale with uncertainty (**Figure 2.3D**, see Methods for details).

When we combine the data across subjects after reflecting of the target directions according to the sign of the mean of the perturbation the generalization data, we observe an asymmetry in the generalization (**Figure 2.3D**). Although we cannot reject the null hypothesis of no effect of uncertainty in the three conditions either in the absolute angle of final hand position ($F_{2,210}$=1.06, p=0.35, two-way repeated measures ANOVA) or in the percent adaptation relative to the learning direction ($F_{2,210}$=0.11, p=0.89, two-way repeated measures ANOVA), the interaction between uncertainty

level and target direction appears to be significant in both in the absolute angle of final hand position ($F_{14,210}$=2.06, p=0.016, two-way repeated measures ANOVA) and in the percent adaptation relative to the learning direction ($F_{14,210}$=1.95, p=0.023, two-way repeated measures ANOVA). The direction that corresponds to maximum generalization (determined by fitting a raised von Mises-like function to each subject and condition data, see Methods for details) is not significantly different from zero for the lower uncertainty conditions (p=0.07 and p=0.27, for σ_p =0° and 4°, respectively; one-sided t-test), but it is significantly different from zero for σ_p =12° (p=0.001, one-sided t-test). Even though it is not consistent with the amounts of uncertainty, there appears to be a weak deviation from a symmetric generalization curve.

One possible explanation for the weak asymmetry that we found is use-dependent learning (Diedrichsen et al., 2010; Huang et al., 2011; Verstynen and Sabes, 2011). Under this hypothesis, subjects will tend to bias their reaching towards highly repeated movements. Hand movements during the testing trials would be attracted to the direction in which the hand moved during learning. To determine whether or not use-dependent learning could account for the observed asymmetry, we first plotted a hypothetical symmetric generalization curve - the angle of final hand position (relative to the angle of the learning target) as a function of target direction (**Figure 2.3E blue dots**) for a perturbation with mean of +30°. Use dependent learning is expected to bias these symmetric movements towards the hand movements during learning (**Figure 2.3E red dots**). It is difficult to quantify this small effect exactly, but we observe that use-dependent learning is consistent with the direction of asymmetry that we see in our data.

To check for more subtle differences in generalization we estimated the width of the generalization curve for each individual subject and uncertainty condition (determined by fitting a raised von Mises-like

function, see Methods for details). For σ_p=0°, 4° and 12° we found generalization widths of 27.0±2.2, 24.0±1.1 and 25.4±1.3 (mean±SEM, across subjects), respectively. We could not conclude that higher uncertainty corresponds to wider generalization for any of the 3 pair-wise comparisons (p=0.79, p=0.13 and p=0.92 for σ_p=12 vs σ_p=0°, σ_p=12° vs σ_p=4° and σ_p=4° vs σ_p=0°, respectively; one-sided paired t-test). These results suggest that the width of generalization of the mean of a noisy visuomotor rotation does not depend on the level of uncertainty in the perturbation.

Finally we did a post-hoc power analysis to compute the minimum detectable effect size (see Methods for details). The rationale behind this kind of analysis is that there may be a difference in generalization widths and that we did not observe it by chance or because the effect size is small over the range of noise levels used here. We computed how big the effect size should be for us to have a high expectation of observing it using a one-sided paired t-test with significance level of 0.05. We determined that we would expect a probability higher than 0.95 of observing a significant difference in the generalization widths, i.e. we would have had sufficient power to detect an effect, if the effect sizes (generalization widths) were higher than 8.5°, 5.7° and 8.2° for the 12° condition relative to 0°, the 12° relative to 4°, and 4° relative to 0°, respectively. Hence we would expect to observe a significant difference in the generalization widths even if the effect size was relatively small.

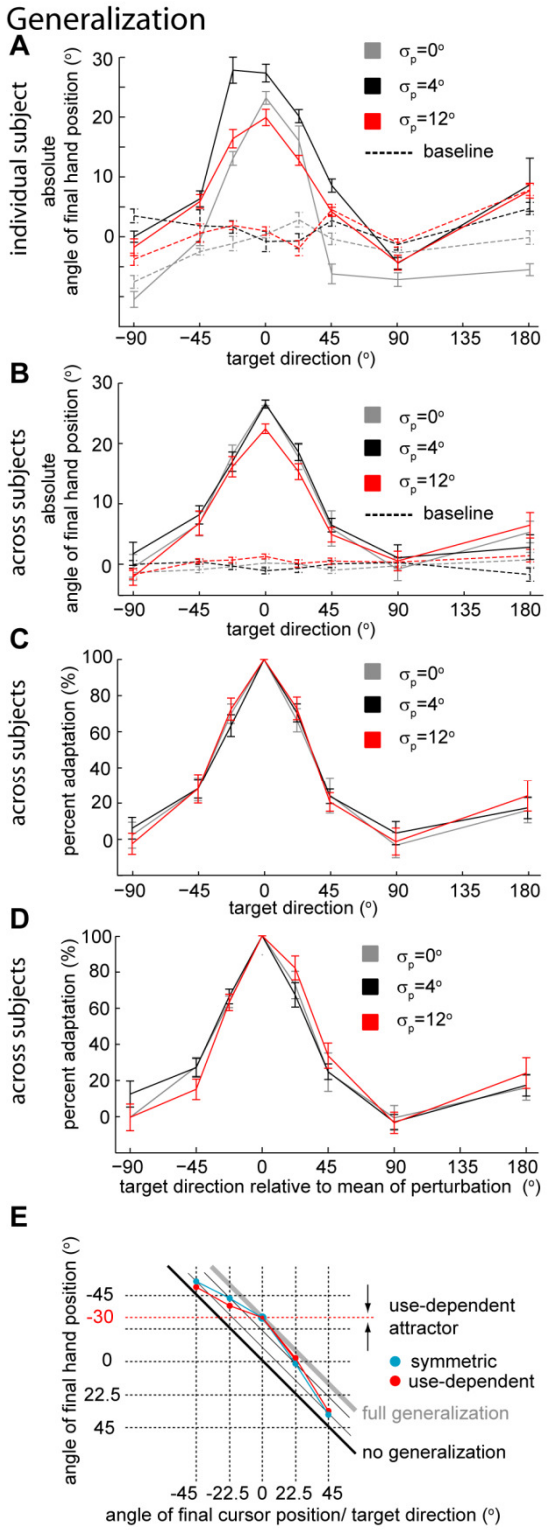

Generalization
A
individual subject
absolute
angle of final hand position (°)
σp=0°
σp=4°
σp=12°
baseline
target direction (°)
B
across subjects
absolute
angle of final hand position (°)
σp=0°
σp=4°
σp=12°
baseline
target direction (°)
C
across subjects
percent adaptation (%)
σp=0°
σp=4°
σp=12°
target direction (°)
D
across subjects
percent adaptation (%)
σp=0°
σp=4°
σp=12°
target direction relative to mean of perturbation (°)
E
angle of final hand position (°)
use-dependent attractor
symmetric
use-dependent
full generalization
no generalization
angle of final cursor position/ target direction (°)

Figure 2.3. Generalization under different variance conditions. A) Baseline and generalization of the mean (±SEM) of a perturbation for a typical subject as measured by the absolute angle of final hand position relative to the target. Solid lines are generalization patterns after learning and dashed lines denote the pre-training (baseline) results. B) Average generalization (±SEM) across subjects. Solid lines denote generalization patterns after learning and dashed lines denote the pre-training (baseline) results. C) Percent adaptation (±SEM) in the generalizing directions relative to the learning direction. D) Percent adaptation (±SEM) in the generalizing directions relative to the learning direction after correcting for the sign of the mean of the perturbation; blocks with -30^{o} mean have the target directions (x-axis) reflected relative to the learning direction. E) Diagram illustrating the direction of an asymmetry caused by used-dependent learning. The blue curve denotes a symmetric, local generalization pattern - without used-dependent learning. If there is used-dependent learning, hand movements in trials towards other targets would be attracted towards the direction to which the hand moved during the learning block (dashed red line). This effect would predict an asymmetry with the same side as the one observed in panel D.

2.4 Discussion

Here we extended traditional movement generalization studies by examining how generalization following learning of a visuomotor rotation is affected by the introduction of trial-by-trial variability. We found that generalization about the mean of a visuomotor rotation is largely unaffected when the perturbation is variable – generalization was local under three different variance conditions. Adaptation is slower and less complete with increased variance level but the width of generalization is unaffected.

We could have expected to see differences in generalization widths. Narrower or broader generalization could both have been justified based on normative arguments or under certain assumptions about the how

uncertainty affects overlapping neural representation of movement. Furthermore, several previous experiments have shown that generalization widths and patterns are neither universally uniform nor immune to changes in experimental conditions. Even though the width of generalization seems to be consistent across tasks such as reaching and wrist tilting (Fernandes et al., 2011; Krakauer et al., 2000), different kinds of perturbations show wider generalization; for example, gain perturbations in center-out reaches appear to generalize globally (Krakauer et al., 2000). Also, studies that manipulate experimental conditions, such as the complexity of the perturbation (Thoroughman and Taylor, 2005) show changes in width of generalization. Moreover, uncertainty has been shown to affect learning (Berniker et al., 2010b; Shea and Kohl, 1990) and retention (Shea and Kohl, 1990), in particular learning of visuomotor rotations (Saijo and Gomi, 2012; Turnham et al., 2012). As uncertainty is important for all of these other aspects of motor learning, it may well affect generalization patterns as well. Here we have shown that generalization width for visuomotor rotations is not affected by changes in variability at least not up to 12 degrees of standard deviation.

A number of models have been proposed for how the nervous system might represent and manipulate probability distributions and uncertainty (Berkes et al., 2011a; Deneve, 2008; Fiser et al., 2010b; Hinton and Sejnowski, 1983b; Hoyer and Hyvarinen, 2003; Ma et al., 2006; Ma, 2010; Sahani and Dayan, 2003; Zemel et al., 1998). Generally in these models, the probability distribution over the set of expected perturbations or other environmental variable based on past experience is called the *prior*. After combining the prior expectations with new incoming sensory information – the *likelihood* - a new probability distribution is computed – the *posterior*. By manipulating the variance of stochastic perturbations we are modifying the variance of the prior and can alter how much subjects rely on new sensory information during single reaches (Körding and Wolpert, 2004).

However, depending on how these distributions are represented by a given neural model and precise assumptions about the neural basis of generalization, these models will make different predictions about generalization behavior.

Some models of neural representation (Girshick et al., 2011; Zemel et al., 1998) explicitly propose an encoding scheme where tuning curves become wider with increasing uncertainty. Under these models neural tuning becomes broader due to the fact that neurons are receiving uncertain input (Zemel et al., 1998) or because they are optimizing the representation of the prior distribution itself with narrowly tuned neurons representing more common directions/orientations (Girshick et al., 2011). Analogous models applied to movement direction would predict that higher uncertainty would lead to broader tuning curves. There is also some experimental data suggesting that individual neurons and populations of neurons are sensitive to changes in uncertainty. Receptive fields in the cat's retina, for instance, become wider with decreasing light levels (Barlow et al., 1957) and populations of neurons in pre-motor cortex appear to be able to represent uncertainty in reach plans (Cisek and Kalaska, 2005). However, there is still relatively limited experimental evidence to constrain these models of the neural representation of uncertainty, particularly in the movement related brain areas. While many electrophysiological experiments have probed how single neurons represent movement-related variables such as hand-direction, speed, or muscle activity (Georgopoulos et al., 1992; Graham et al., 2003; Kakei et al., 1999; Moran and Schwartz, 1999; Sergio et al., 2005) and even how neural responses change during adaptation to visuomotor rotations (Paz et al., 2003), relatively little is known about how neural activity changes in the presence of sensorimotor uncertainty (but see Britten et al., 1992; Cisek and Kalaska, 2002, 2005; Rickert et al., 2009).

If it is true that the width of generalization curves reflects the tuning widths of the neurons, we did not find signs of such broadening in our

generalization study. Importantly, there are three natural interpretations of this result. It could be that our study failed to see the effect because we did not have the necessary statistical power. However, with 16 subjects we ran far more subjects than most movement studies. Also, our power analysis revealed that we should have seen even relatively small effects of broadening; therefore it seems unlikely that this effect exists and we were unable to observe it. Another possibility is that theories that predict broadening of tuning curves are wrong, or at least do not apply to simple targeted reaching movements. However, none of the theories that deal with the representation of uncertainty explicitly mention their predictions of generalization and (third interpretation) generalization may be related to underlying neural representations in a more complex way than generally assumed in motor control research (Donchin et al., 2003; Krakauer et al., 2000a; Thoroughman and Shadmehr, 2000; Thoroughman and Taylor, 2005).

We have found weak signs that generalization curves are slightly asymmetric. Use-dependent learning, where subjects are biased to move in a way that is similar to how they have been moving previously is one of the newly emerging insights in computational motor control (Diedrichsen et al., 2010; Huang et al., 2011; Verstynen and Sabes, 2011). These theories would suggest biases towards the typical direction of hand movement. We find that this is consistent with the weak asymmetry that we found in the generalization curves. Furthermore it is also a potential explanation for the commonly observed adaptation at 180^{o} (Donchin et al., 2003; Fernandes et al., 2011; Krakauer et al., 2000), since movements in this direction are similar to movements returning from the learning target to the center of the working space. Future research would be necessary to clarify which factors give rise to this asymmetry. For example, this asymmetry may disappear if perturbations are introduced in a gradual manner or if limb mechanics are controlled in more detail.

For some subjects the simplicity of the task and the salience of the perturbations led to cognitive strategies that may have introduced noise in the measurements. As such, we found relatively high variability across subjects. Gradually introduced perturbations have been shown to lead to a more complete adaptation and larger aftereffects (Kagerer et al., 1997; Taylor and Ivry, 2011b; Turnham et al., 2012). It would be interesting to test if slowly introduced perturbations would reduce the subject-by-subject variance and even have some effect in the generalization widths.

The focus both in behavioral as well as in electrophysiological studies in motor control has been on the generalization and representation of perturbations without any trial-by-trial variability. While uncertainty has been shown to be important in many behavioral settings, variability does not appear to change generalization curves during visuomotor rotation. Variability does affect learning, however, and understanding how variability affects generalization in other tasks should provide some insight into the neural representations of uncertainty and movement.

2.5 Materials and Methods

Ethics statement. The experimental protocol was approved by the Northwestern University Institutional Review Board and is in accordance with the Northwestern University Institutional Review Board's policy statement on the use of human subjects in experiments. Written informed consent was obtained from all participants. The Institutional Review Board of Northwestern University approved the study.

Experimental protocol. Sixteen right-handed healthy subjects (5 male, 11 female; aged 27±3.2 years) participated in the experiment. All were naive to the purpose of the experiment, and were paid according to their performance. Subjects made center-out reaches in an approximately 150 x 150mm central region of a 100cm x 70cm workspace. They controlled the position of a cursor with their right index finger, which was recorded using an Optotrak 3D Investigator Motion Capture System. A projector and mirror system was calibrated such that visual feedback was perceived as being in the movement plane (**Figure 2.1A**), and the subject's view of their hand was blocked by the mirror.

The task was designed to measure how subjects generalize the mean of a noisy visuomotor rotation, that is, how a perturbation learned during movements in one direction affects subsequent movements in other, test directions. This experiment extends a previous paradigm that allows measurement of generalization about a fixed perturbation (Krakauer et al., 2000) to include stochastic perturbations.

Subjects were instructed to make center-out reaches into a certain direction (the learning direction) until they adapted to the perturbations/rotations. During this period subjects were given endpoint feedback - that is, the final position of the hidden cursor was displayed - and were eventually able to correct endpoint errors in the learning direction. Afterwards, they were instructed to make movements into other directions (the generalizing directions) in order to measure the generalization pattern of the learned mean of the perturbation. Generalization of the mean was assessed by analyzing their average reaching direction for each target.

The learning direction was randomly sampled from one of the 4 diagonal directions and generalization was measured in 7 directions: 180^{o}, $\pm 90^{o}$, $\pm 45^{o}$ and $\pm 25^{o}$ from the learning direction (**Figure 2.1C**). Subjects controlled the position of a red circle, the cursor (~3mm radius), with their

right index finger. Except for the first familiarization trials the position of the cursor was hidden. Subjects were instructed to make radial reaches from a central blue circle, the starting circle (~6mm radius) to one of 8 yellow circles, the targets (~6mm radius). Targets were all displayed at a distance of 72mm from the central blue circle. 300ms after positioning the cursor over the blue circle, the cursor disappeared, one of the eight targets appeared and subjects had to reach it. On some of the trials the final position of the cursor was displayed for 500ms (endpoint feedback). The final position of the cursor was defined as the first position of the cursor when its center was at a distance greater than 72mm from the center of the starting circle. If the reach was successful, that is, if the center of the red cursor was inside the target then the target turned white and subjects were rewarded by having a point added to their score. If a successful reach happened in those trials where no information was provided about the success of the reach (no endpoint feedback) then a point was added to a hidden score. To initiate the next trial, subjects had to reposition the cursor in the starting blue circle. Except for the familiarization trials where the cursor was always visible, the cursor was visible only within a distance of 10mm from the center of the starting blue circle. Since some subjects have difficulty finding their way back to the starting blue circle, 4 seconds after the previous trial was over, the cursor flashed every second for 50ms to allow subjects to find the starting position.

We measured generalization of the learned mean for a rotation of $\pm 30^{o}$ under three variability levels. Each trial, noise was added to the visuomotor rotation drawn from a Gaussian with a standard deviation of 0^{o}, 4^{o} or 12^{o}. The standard deviation of 0^{o} reproduces previous experiments that measured the generalization pattern of a deterministic visuomotor rotation (Krakauer et al., 2000).

The experiment was divided into three blocks of 560 trials (**Figure 2.1B)**. Blocks differed in the level of variance and were pseudo-randomized.

Each block was composed of 4 sub-blocks: *Familiarization*, *Baseline*, *Learning* and *Testing*. No rotation was imposed during the familiarization and baseline blocks. In all cases, the maximum time to complete each trial was 4 seconds and the minimum time 40ms. If any of these times was violated the trial was restarted.

Familiarization. During the first half (40 trials, 5 movements to each target) of the familiarization sub-block the cursor was always visible. During the second half (40 trials, 5 movements to each target) only endpoint position was displayed.

Baseline. This sub-block was used to measure the baseline (80 trials, 10 movements to each target). These reaches were made under the same conditions as the second half of the familiarization block – endpoint feedback was provided in all trials and no perturbation was applied to the cursor.

Learning. Subjects completed 240 trials of movements towards a single learning target with only endpoint feedback. The cursor was rotated relative to hand position.

Testing. The testing sub-block was composed of 160 trials. In order to prevent de-adaptation to the perturbation, the learning target was revisited at least twice every 4 trials; every sequence of 4 trials consisted of two reaches towards the learning target and two reaches towards any two of the 8 targets. Targets were chosen pseudo-randomly so that there were 10 reaches total towards each of the generalizing directions. Endpoint feedback is provided only in the learning direction trials. During these trials towards the learning direction the perturbations applied to the cursor position were sampled from the same distribution used in the learning block.

Data analysis. Final hand position angle gives us a measure of the subject's estimation of the perturbation. For each trial we computed the final hand position by averaging the last data point before the hand goes beyond a distance of 72mm – the target radial distance – from the center of the starting circle and the first data point after that. Notice that final hand position is well defined for every trial since trials were restarted whenever the subject did not go beyond a distance of 72mm.

Absolute final hand position and percent adaptation. Since the sign of the mean of the distribution of perturbation was randomly chosen for each block and each subject, we normalized the angle of final hand position according to the sign of mean of the perturbations so that the average final hand position angle in the learning direction was positive for every block; this was done by multiplying by -1 the angle of final hand position when the mean of the distribution of perturbations was positive (+30 degrees). We call this measure the *absolute final hand position.* We measured the baseline movement biases, $b(\theta)$, and the learned and generalized means, $g(\theta)$, by considering the average absolute angle of final hand position (**Figure 2.2**). Specifically, $b(\theta_t) = \theta_t - \overline{\theta}^{b}_{h,t}$ and $g(\theta_t) = \theta_t - \overline{\theta}^{g}_{h,t}$, where θ_t is target direction, $\overline{\theta}^{b,t}_{h}$ and $\overline{\theta}^{g,t}_{h}$ are average absolute angles of final hand position in trials towards target t during baseline and testing, respectively. Using this information we can compute the *percent adaptation*, that is, the difference between testing and baseline in the each direction relative to the learning direction θ_l (**Figure 2.3C**):

$$\text{percent adaptation}(\theta_t) = \frac{g(\theta_t) - b(\theta_t)}{g(\theta_l) - b(\theta_l)} \times 100$$

Notice that a positive absolute angle of final hand position or percent adaptation corresponds to a hand movement that counteracts the mean of

the distribution of perturbations. We use one of these two measures in every figure and analysis (with the exception of **Figure 2.1A** and **2.3E** where the true sign of final hand angle is displayed).

Time-scales of learning. To compute the time scales and amount of adaptation we considered only the first block of learning for each subject (n=8, n=4 and n=4 for σ_p =0°, 4° and 12°, respectively). We then fitted exponential learning curves that were constrained to start at zero. We used bootstrapping over trials to determine the p-value for the differences between the timescales of learning and between adaptation at end of the learning sub-blocks.

Correcting for the sign of the mean of the perturbation. For part of the analysis (**Figure 2.3D**) we wanted to take into account the fact that, for some of the blocks, the mean of the imposed perturbation had negative sign (-30°). This was done with the objective of searching for aspects of generalization that could depend on the sign of the imposed perturbation. We did the correction by reflecting the target directions relative to the learning target direction; if we set the learning target direction, θ_l to be zero, then the corrected generalization function $g^c(\theta)$ is defined as: $g^c(\theta) = g(-\theta)$.

Width of generalization. To determine the generalization width we used raised von Mises-like (circular Gaussian) functions:

$$g(\theta \mid \beta_0, \beta_1, \beta_2, \beta_3) = b_0 + b_1 \exp^{\beta_2 \cos(\theta - \beta_3)} \qquad 2.1$$

where θ is target direction. We fitted these functions to each individual percent adaptation generalization. We used $1/\sqrt{\beta_2}$ as the estimate of generalization width. We excluded two subjects from this analysis because

the estimated width of their generalization in at least one of the uncertainty conditions was more than 10 standard deviations away from the mean of the remaining subjects' widths for that uncertainty condition.

Peak of generalization. To determine if there is a consistent asymmetry in the generalization pattern, we determined, for each subject and each uncertainty condition, the angle of maximum generalization given by the parameter β_3 in **Equation 2.1**. The sign of the parameter was corrected for the sign of the mean of the perturbation, more specifically, we multiplied β_3 by the sign of the mean of the perturbation.

Effect size. To compute the minimum effect size, η, that would have been required for detecting an significant effect with probability above 0.95 using a two-sample one-sided paired t-test at a significance level of 0.05, we used the standard minimum detectable effect formula (e.g. see Zar, 1999)

$$\eta = \sqrt{\frac{s_1^2 + s_2^2}{14}}(t_{0.05,26} + t_{0.05,26})$$

where s_1 and s_2 are the estimated variances of widths for each uncertainty condition and $t_{\alpha,\nu}$ represents the value of the inverse of the cumulative t-student distribution with ν degrees of freedom at $1-\alpha$.

3. The Generalization of Prior Uncertainty During Reaching

Hugo L. Fernandes, Ian H. Stevenson and Konrad P. Kording

Article under review

Author Contributions. Conceived and designed the experiments: HLF IHS KPK. Performed the experiments: HLF. Analyzed the data: HLF IHS KPK. Implemented the analysis: HLF with the help of IHS. Wrote the paper: HLF IHS KPK

3.1 Summary

Bayesian statistics defines how new information, given by a likelihood, should be combined with previously acquired information, given by a prior distribution. Many experiments have shown that humans make use of such priors in cognitive, perceptual, and motor tasks, but where do priors come from? As people never experience the same situation twice, they can only construct priors by generalizing from similar past experiences. Here we examine the generalization of priors over stochastic visuomotor perturbations in reaching experiments. In particular, we look into how the first two moments of the prior - the mean and variance (uncertainty) - generalize. We find that uncertainty appears to generalize differently from the mean of the prior, and an interesting asymmetry arises when the mean and the uncertainty are manipulated simultaneously.

3.2 Introduction

People use priors during sensorimotor tasks, and such priors allow perception and movement to be more accurate in many situations (Alais and Burr, 2004; Jazayeri and Shadlen, 2009; Körding and Wolpert, 2004a; Tassinari et al., 2006). In Bayesian statistics the prior reflects information accumulated from previous experience, which is then combined with incoming sensory feedback (the likelihood). As we interact with the world, we learn about its statistics (e.g. means and variances) and incorporate this information into our priors. However, since we are never in the same situation twice, we must use past information from different but similar situations to derive the right prior beliefs for a specific task. Only by generalizing from past situations to our current one can we calculate what to expect.

In asking how humans generalize priors it is essential to understand how we represent uncertainty. There are a number of models of how the nervous system might represent uncertainty (Fiser et al., 2010a; Hoyer and Hyvärinen, 2003; Ma et al., 2006). However, there is limited experimental evidence to constrain these models. Many electrophysiological experiments have probed how single neurons represent movement-related variables such as hand-direction, speed, or muscle activity (Georgopoulos et al., 1992; Kakei et al., 1999; Moran and Schwartz, 1999; Sergio et al., 2005), but relatively little is known about the neural representation of uncertainty in sensorimotor tasks (Britten et al., 1992; Cisek and Kalaska, 2005; Rickert et al., 2009). Furthermore, to our knowledge, none of the theoretical models for neural representations of uncertainty makes any prediction for generalization of priors nor is there an established normative conjecture for how behaviors *should* generalize.

One way of characterizing the generalization of priors comes from previous generalization experiments in motor control (Donchin et al., 2003;

Ghahramani et al., 1996a; Mattar and Ostry, 2007; Shadmehr, 2004; Thoroughman and Shadmehr, 2000). During center-out reaching, training with a rotational perturbation in one direction biases movements to nearby targets, and this bias decreases with increasing distance from the training direction (Krakauer et al., 2000). Previous studies have looked into whether uncertainty affects this generalization pattern (Fernandes et al., 2012). However how uncertainty itself might generalize is unknown.

Here, we manipulated the mean and the variance (uncertainty) of a noisy visuomotor rotation (the prior) imposed during movements in one (training) direction. After training, we examined subjects' movements in test directions and measured subjects' uncertainty by probing their reliance on feedback (the likelihood) (Körding and Wolpert, 2004). In a *first experiment* we manipulated the variance without changing the mean. As with standard rotational generalization, we found a strong local effect where subjects' uncertainty peaks in the training direction and decreases with increasing distance. However, unlike standard rotational generalization, we found that changes in uncertainty had a global effect. In a *second experiment* we manipulated the variance while introducing a mean perturbation and observed interesting nonlinear interactions between mean and variance -- subjects had the highest uncertainty not in the training direction but in a direction away from the perturbation.

3.3 Materials and Methods

Ethics statement. The study and all experimental protocols were approved by the Northwestern University Institutional Review Board and are in accordance with the Northwestern University Institutional Review Board's

policy statement on the use of human subjects in experiments. Written informed consent was obtained from all participants.

Experimental protocol. General. Forty right-handed healthy subjects (15 male, 25 female; aged 28.5±3.5 years) participated in the experiments; n=32 in Experiment 1 and n=8 in Experiment 2. All were naive to the purpose of the experiments, and were paid according to their performance. Subjects made center-out reaches in a 150 x 150 mm workspace. They controlled the position of a cursor with their right index finger, which was recorded using an Optotrak 3D Investigator Motion Capture System. A projector and mirror system was calibrated such that visual feedback was perceived as being in the movement plane (**Figure 3.1A** and Fernandes et al., 2012), and the subject's view of their hand was blocked by the mirror.

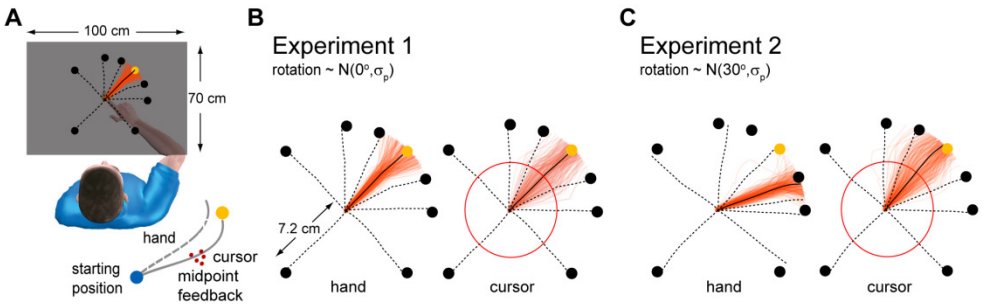


Figure 3.1. Experimental setup and typical trajectory data. A, Subjects move a hidden cursor from a starting position to a target (yellow) by moving their occluded right index finger. We measure the generalization of the learned variance of a perturbation utilizing the response to a noisy midpoint cursor feedback (red dots). B, Experiment 1, with zero mean perturbation. Subject's hand and cursor position during learning trials (red trials, black average). Average trajectories for generalizing directions are shown as black dashed lines (corresponding targets are black dots).

The position where midpoint feedback is triggered is denoted by the red line. C, Experiment 2, where the absolute mean perturbation is 30°. Same notation as (B).

The task was designed to measure how subjects generalize a learned variance (and the learned mean) of a noisy visuomotor rotation, that is, how the uncertainty related to a perturbation learned during movements into one direction affects subsequent movements into other, test directions. The experiments combined two previously existing paradigms; one that allows measurement of generalization of the mean of perturbations (Fernandes et al., 2012; Krakauer et al., 2000) and another that allows measurement of how uncertain subjects are of a perturbation (Körding and Wolpert, 2004).

Subjects were instructed to make reaches into a certain direction (the *learning direction*) until they adapted to the perturbations. During this period subjects were given endpoint feedback, and were eventually able to correct endpoint errors in the learning direction. Afterwards, they were instructed to make movements into other directions (the *generalizing directions*) in order to measure the generalization patterns of the learned mean and of the learned variance of the perturbation. Generalization patterns were assessed by using the fact that subjects combine their previous knowledge about the distribution of perturbations (the prior) and the feedback (the likelihood, see below) that they receive midway through the movement about the true position of the cursor weighted by their relative uncertainties (Körding and Wolpert, 2004). The ideal way to combine these two sources of information is to combine the means of the prior and likelihood, weighted by their relative precision (the inverse of the variance). Assuming that subjects combine this information optimally we can measure their relative uncertainty by computing the slope of a linear regression of the negative of final hand position (subjects' estimated perturbation) as a

function of the perturbation (see **Equations 3.1** and **3.2** and **Figure 3.2A**). Analogously we are able to simultaneously measure the mean of the prior (see **Equation 3.3**).

In order to probe uncertainty, 5 red circles identical to the cursor are flashed midway through every trial reach: the *midpoint feedback* (likelihood). The position of these dots is sampled from an isotropic two-dimensional Normal distribution centered on the true position of the cursor with variance ~5.1mm (chosen empirically to avoid complete reliance on either prior or likelihood, see below). Hence they give uncertain information about the true position of the cursor. The midpoint feedback is shown already during the familiarization block. This way subjects get a better idea of how the dots relate to the position of the cursor. We use the final hand position to measure the level of uncertainty that the subject has on the hidden cursor position. Describing it within the nomenclature of the Bayesian framework, the perturbation is sampled from a distribution with defined mean and variance and which is approximated by the prior, the midway flashing dots that give uncertain information about the true position of the red cursor correspond to the likelihood and the estimated perturbation corresponds to the mean of the posterior. By looking at the slope of the negative of the final hand position (mean of the posterior) as a linear function of the perturbation (**Figure 3.2A**) we can estimate the relative reliance on prior information -- relative to midpoint feedback information (Körding and Wolpert, 2004). Hence we can compute a relative measure of subjects' learned and generalized uncertainty.

The learning direction was randomly sampled from one of the 4 diagonal directions and generalization was measured in 7 directions displayed at 180^o, ±90^o, ±45^o and ±25^o degrees from the learning direction. Subjects control the position of a red circle, the cursor (~3mm radius), with their right index finger. Except for the first familiarization trials the position of the cursor is hidden. Subjects were instructed to make radial reaches from a

central blue circle, the starting circle (~6mm radius) to one of 8 yellow circles, the targets (~6mm radius). Targets were all displayed at a distance of 72mm from the central blue circle. 300ms after positioning the cursor over the blue circle, the cursor disappeared, one of the eight targets appeared and subjects had to reach it. On some of the trials the final position of the cursor was displayed for 500ms (endpoint feedback). If the reach was successful, that is, if the center of the red cursor was inside the target then the target turned white and subjects were rewarded by having a point added to their score. If a successful reach happened in those trials where no information was provided about the success of the reach (no endpoint feedback) then a point was added to a hidden score. To begin the next trial, subjects had to reposition the cursor in the starting blue circle. Except for the familiarization trials where the cursor was always visible, the cursor was visible only within a distance of 10mm from the center of the starting blue circle. Since some subjects have difficulty finding their way back to the starting blue circle, 4 seconds after the previous trial was over, the cursor flashed every second for 50ms to allow subjects to find the starting position.

The study comprised two experiments; Experiments 1 and 2. The experiments differ in that the mean of the imposed perturbations is zero in Experiment 1 and nonzero in Experiment 2.

Experiment 1: Generalization of uncertainty under zero mean rotation. The goal of Experiment 1 is to measure the generalization pattern of uncertainty. The experiment begins with an initial *Familiarization* block (40 trials, 5 movements to each target) where the cursor is always visible. No rotation was imposed during the familiarization block. After that, the experiment is divided into two blocks of 720 trials, one for each level of variability (std: 4^o or 12^o). The two blocks differ only in level of variance and their order is pseudo-randomized across subjects.

Each block of 720 trials is composed by a *Learning* and a *Testing* sub-blocks. The learning direction is the same for both blocks, but selected randomly for each subject from 4 possible directions; ±45° and ±135°. The maximum time to complete each trial is 4 seconds and there is a minimum time of 400ms to complete the second half of the reach. If any of these times is violated the trial is restarted. The minimum time threshold is to guarantee that subjects have enough time to integrate the midpoint feedback information.

Learning. Subjects complete 240 trials of movements towards a single learning target with midpoint (the cloud of circles flashed midway through the movement) and endpoint feedback. The cursor is hidden and rotated relative to hand position. The rotations applied within each block are sampled from the same normal distribution with mean 0° and standard deviation pseudo-randomly chosen to be either 4° or 12°.

Testing Uncertainty. The testing uncertainty sub-block (480 trials) is composed by sequences of 4 trials. In order to prevent forgetting of the perturbation the first 2 trials of these sequences are towards the learning direction and the other 2 towards any 2 of the 8 possible directions. Targets are chosen pseudo-randomly so that exactly 20 reaches are made towards each generalization target. Endpoint feedback is provided only in trials towards the learning direction and midpoint feedback is provided in all directions.

Experiment 2: Generalization of uncertainty under nonzero mean rotation. Experiment 2 is aimed at measuring the generalization pattern of mean and variance when both are perturbed simultaneously. The purpose of Experiment 2 is to distinguish between the abstract Bayesian models that explain the data of Experiment 1 and to see how changing the mean of a perturbation influences the generalization of uncertainty. Hence, the

difference between Experiments 1 and 2 is essentially that in Experiment 2 the perturbations have a nonzero mean. The experiment starts with a *Familiarization* block (40 trials) just like the one in Experiment 1, and it is then divided into two blocks of 880 trials. The reason for the larger number of trials is that there is an extra sub-block between the *Learning* (240 trials) and the *Testing Uncertainty* (480 trials) sub-blocks; the *Testing Mean* sub-block (160 trials).

Testing Mean. Sub-block for measuring the generalization of the mean (160 trials). This sub-block allows us to measure directly how each subject generalized the mean of the perturbation (see **Figure 3.5A, B**). Subjects make reaches towards all targets. Endpoint feedback and midpoint feedback are provided only in the learning direction trials. In order to prevent de-adaptation to the perturbation, the learning target is revisited at least twice every 4 trials; every sequence of 4 trials is composed by two reaches towards the learning target and two reaches towards any two of the 8 targets. Targets are chosen pseudo-randomly so that there are in total 10 reaches towards each of the generalizing directions. Even though midpoint feedback is not displayed in movements towards generalizing direction, there is still a minimum amount of time to complete the second part of the movement in every trial. Hence, subjects still slow down halfway through the movement as in the trials where midpoint feedback is displayed.

We can then use these measurements of the generalization of the mean during the *Testing Uncertainty* sub-block; in each target direction, the perturbation will have a mean equal to how much the subject generalized the learned mean to that direction (as measured in the *Testing Mean* sub-block, see **Figure 3.5A, B**). Notice that, even though there is no endpoint feedback during *Testing Uncertainty*, if the mean perturbation doesn't match the subject's generalized mean then the midpoint feedback could perturb subjects learned mean and uncertainty, and, consequently the measurement of generalization of uncertainty. Hence, by matching the

mean of the probing perturbation in each of the generalizing directions with the learned mean, we minimize the possibility of subjects learning from the midpoint feedback information (see description of the experimental design below for further details regarding this issue).

Experimental protocol. Details. There are two important details to be noticed regarding the experimental design of both experiments.

It is not possible to measure a baseline for uncertainty using this protocol. It is not possible to measure a baseline for relative reliance on midpoint feedback due to the fact that we need to introduce a perturbation to measure the slope (**Equation 3.1** and **3.2** and **Figure 3.2A**). For that reason we measured the generalization of two different levels of variability – standard deviation of 4^o and 12^o. These standard deviation values were chosen empirically based on several constrains that the task imposes: at the same time that both values need to be sufficiently different, the smaller variance cannot be too small otherwise the range of the perturbations is not large enough to measure the relative reliance on midpoint feedback (the slope of a linear regression) with a reasonable confidence interval. The higher variance condition cannot be too large otherwise it could introduce nonlinearities (Körding et al., 2007; Wei and Kording, 2009) and because of the constrains inherent of working in a circular support. The standard deviation of the likelihood was chosen empirically so that the slopes would be close to 0.5. This is the range where behavior is influenced equally by prior and likelihood, and thus where fluctuations in uncertainty will have the most effect. Several values were tested while designing the experiment, starting with the theoretical value that would produce the desired slope and changing it until values obtained for the slope were around 0.5.

In the generalizing directions, the standard deviation of the perturbation used to probe uncertainty is the same regardless of the standard deviation

of the imposed perturbation in the learning direction. Since midpoint feedback is necessary to measure subject's relative uncertainty, special care is needed to ensure that this feedback does not bias measurements of generalization. Differences in learning and sensorimotor integration could both lead to spurious differences in the patterns of relative uncertainty. In both experiments we do not provide endpoint feedback in the generalizing directions. In Experiment 1 this is enough to ensure that differences in generalization patterns cannot be due to learning during the testing phase. However, during both experiments, perturbed midpoint feedback is the only method to measure each subject's relative uncertainty. The spread and timing of the midpoint feedback was the same across the two variance conditions. Additionally, we set the variance of the perturbation in the generalization directions to the geometric mean of the two standard deviations used in the learning directions, namely $\sqrt{4 \times 12}$ °. This guarantees that the only difference in the distribution of perturbations between the blocks of trials during movements in the learning direction. The important consequence is that, even if the midpoint feedback allowed subjects to learn during generalization trials, learning would only act to bring the two generalizations curves closer together. The methods used here, thus, set a lower bound on the distance between the generalization patterns for the two variance conditions.

Data analysis. General.

Final hand position and estimated perturbation. In this paradigm, the final hand position angle, θ_{fh} angle gives us a measure of subjects estimated perturbation $\hat{\theta}$, specifically; $\hat{\theta} = -\theta_{\text{fh}}$. We compute the final hand position for each trial by averaging the last data point before the hand goes beyond a distance of 72mm – the target distance – from the center of the starting

circle and the first data point after that. Every trial was restarted if subjects didn't go beyond the target distance, thus θ_{fh} and $\hat{\theta}$ are defined for every trial.

Measuring generalization of uncertainty; relative reliance on midpoint feedback (slope). We assume that the estimated perturbation corresponds to the mean of the posterior (Körding and Wolpert, 2004). Assuming Gaussian distributions, an ideal observer/actor would combine information from their prior over cursor perturbations (θ_p) and the perturbation angle sensed from the midpoint feedback information (θ_f) weighting their values by their relative precisions, according to

$$\hat{\theta} = \frac{\sigma_f^2}{\sigma_p^2 + \sigma_f^2}\theta_p + \frac{\sigma_p^2}{\sigma_p^2 + \sigma_f^2}\theta_f \qquad 3.1$$

Where σ_p^2 and σ_f^2 denote subjects' uncertainty in prior and midpoint feedback respectively. Subjects estimated angle of the perturbation ($\hat{\theta}$) is reflected in the angle of their final hand position. As a proxy for the sensed perturbation angle (θ_f) we use the real perturbation angle corrected (see below) for the bias in the centroid of the sampled flashing dots of the likelihood. That is, instead of considering the real perturbation angle we considered the angle that a vector from the center of the central blue circle to the centroid of the flashing dots would do with the target direction if the subject had moved straight to the target. Importantly, this equation allows us to estimate the relative learned variance of the prior for each generalizing direction, and to compute a relative generalization function for uncertainty. We estimate the value of the slope (s_p)

$$s_p = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_f^2} \qquad 3.2$$

for each variance condition and direction using linear regression (**Figure 3.2**A) – we use the median of bootstrap samples to reduce the influence of outliers when computing the slopes. This slope, the relative uncertainty, serves as the basis for most of our analysis. The centroid adjustment and bootstrapped slope estimates provide more robust measures of behavior, but using unadjusted perturbations and maximum likelihood estimated slope produce qualitatively very similar results.

Measuring generalization of the mean.

Inferred mean. We are able to infer the mean of the prior in both experiments using the data from the *testing uncertainty* block. We do this by computing the intercept of a linear regression; we can rearrange Equation 3.1 to obtain

$$\theta_f = \theta_p + \frac{\sigma_f^2 + \sigma_p^2}{\sigma_f^2}(\theta_f - \hat{\theta}) \qquad 3.3$$

We can hence, for each target direction, use as estimate of the subjects' mean of the prior, θ_p, the intercept of linear regression of θ_f as a function of $\theta_f - \hat{\theta}$.

Direct measurement of the prior's mean. In Experiment 2, during the *testing mean* block, we were able to directly measure generalization of the mean in each of the generalizing direction; during the trials in the generalizing directions of the *testing mean* sub-block, subjects were not shown midpoint feedback (the likelihood) and hence their estimate - as inferred by final hand

position - is assumed to be the mean of the prior distribution. Using this information we were able to compute, during the experiment, the means of the perturbations used to probe uncertainty in the generalizing directions during the *testing uncertainty* sub-block (see **Experimental protocol** above for details). During trials in the learning direction subjects were still shown midpoint feedback. Thus, their average final hand position during trials towards the learning direction is an estimate of the mean of the posterior and not of the prior. The generalization patterns obtained during the test of the mean block match very well the ones inferred using the data from the testing of uncertainty blocks ($F_{1,7}$=3.27, p=0.11, two-way (testing block, direction) repeated measures (subject) ANOVA, see **Figure 3.5**B; see also **Figure 3.5**A first and second rows for individual subject data). The higher complexity of this task, relative to previous studies that measured generalization of means (Fernandes et al., 2012), lead to smaller variability (possibly due to smaller variability in cognitive strategies) across subjects.

Absolute mean and percent adaptation. In Experiment 2, since for half of the subjects the mean of the perturbation was -30°, we normalized the estimated perturbation (as measured by the negative of the angle of final hand position) according to the sign of the mean of the perturbation; we multiplied by -1 the angle of the estimated perturbation if the mean of the perturbation was negative (-30°). Hence, a positive absolute mean corresponds to a movement that counteracts the perturbation. We call the measurements of the mean using this normalization the *inferred absolute mean* and the *measured absolute mean*. Using the *absolute inferred mean* we compute the *percent adaptation*, the amount of learned/generalized mean relative to the learned mean in the learning direction (**Figure 3.4**E). The percent adaptation in the learning direction is hence, by definition, 100%.

Models for generalization.

The models we consider are online learning models. They use gradient descent to learn the mean and standard deviation of the imposed prior; for each trial they use gradient descent to minimize the expected squared error between the target angle and the final cursor position angle.

Consider the subjects' original prior:

$$p_0 = \mathcal{N}(\theta_0, \sigma_0) = \mathcal{N}(0^\circ, \sigma_0)$$

.

We assume that the original prior is the same for all directions. Throughout this section we generally use θ to denote perturbation related angles and ϕ to denote target angles. Note that the perturbation angles θ are always given relative to a target direction.

We define a context function, $W_{\phi_l}(\phi_g)$, for a target direction ϕ_g relative to the learning target direction ϕ_l, as a scaled von Mises function:

$$W_{\phi_l}(\phi_g) = \left(b_0 + \exp\left(b_1 \cos\left(\phi_g - \phi_l\right)\right)\right) / \alpha$$

where b_0 is a baseline for context, b_1 defines the width of the context and $\alpha = b_0 + \exp(b_1)$ is a normalization factor so that context is 1 in the learning direction, that is, $W_{\phi_l}(\phi_l) = 1$. This is the same as saying that generalization is complete in the learning direction. The context function can be interpreted as defining how similar the movement directions.

The Model's parameters are: three context function parameters allowing for different context baselines for mean and variance, $b_0^{\theta_p}$, $b_0^{\sigma_p}$, b_1; the initial prior uncertainty σ_0 and likelihood uncertainty σ_f and the learning rates of mean and variance, η_{θ_p} and η_{σ_p}. These 7 parameters were enough to produce good fits to the data from the first block. However, we observed that, none of the models managed to capture decreases in the variance of the prior during the second block. For this reason and to account for possible differences between the learning and the testing blocks, we added an extra parameter that scales the model's output of prior uncertainty at the end of learning before fitting it to the testing data. Hence both Models 1 and 2 have a total of 8 parameters. While Model 1 is the target centered, Model 2 tests the hypothesis that generalization of variance has a non-target centered reference frame; the visual feedback information.

Model 1. Gradient descent with target centered reference frame. On each i-th trial of the learning block, the subject is trying to minimize the squared error between the target angle and the final cursor position angle, that is, trying to learn the mean $\hat{\theta}_p$ and variance $\hat{\sigma}_p^2$ of the prior imposed in the learning direction such that

$$(\hat{\theta}_p^{\phi_l}, \hat{\sigma}_p^{\phi_l}) = \arg\min_{\theta_p, \sigma_p} e\left(\theta^i, \theta_p, \sigma_p\right)$$

where θ_i is the perturbation imposed during the i-th trial and $e\left(\theta^i, \theta_p, \sigma_p\right) = \left(\theta^i - \hat{\theta}\left(\theta_p, \sigma_p\right)\right)^2$ where $\hat{\theta}\left(\theta_p, \sigma_p\right)$ is defined in Equation 3.1. The model takes as input the learning trials and assumes that the standard deviation and mean of the subject's prior evolve according to

$$\sigma_p^{\phi_g,i} = \sigma_p^{\phi_g,i-1} - \eta_{\sigma_p} W_{\phi_l}(\phi_g) \left.\frac{\partial e}{\partial \sigma_p}\right|_{\theta_p^{\phi_l,i-1},\sigma_p^{\phi_l,i-1}}$$

$$\theta_p^{\phi_g,i} = \theta_p^{\phi_g,i-1} - \eta_{\theta_p} W_{\phi_l}(\phi_g) \left.\frac{\partial e}{\partial \theta_p}\right|_{\theta_p^{\phi_l,i-1},\sigma_p^{\phi_l,i-1}}$$

Model 2. Gradient descent with difference frames of reference; visual feedback centered context for generalization for variance. The only difference between the models is that the context function for uncertainty in the prior (standard deviation, σ_p) is centered on the angle of the centroid of the displayed cloud of dots, ϕ_c, while the context function for the mean remains centered on the learning target direction, ϕ_l:

$$\sigma_p^{\phi_g,i} = \sigma_p^{\phi_g,i-1} - \eta_{\sigma_p} W_{\phi_c}(\phi_g) \left.\frac{\partial e}{\partial \sigma_p}\right|_{\theta_p^{\phi_l,i-1},\sigma_p^{\phi_c,i-1}}$$

$$\mu_p^{\phi_g,i} = \theta_p^{\phi_g,i-1} - \eta_{\theta_p} W_{\phi_l}(\phi_g) \left.\frac{\partial e}{\partial \theta_p}\right|_{\theta_p^{\phi_l,i-1},\sigma_p^{\phi_c,i-1}}$$

In order to avoid using behavioral data obtained during the learning block, the model uses the predicted position of the cloud of dots as a proxy for ϕ_c. This predicted position is obtained by computing, give the trial perturbation, where the cloud of dots would appear if the subject performed a straight center-out movement corrected by the current mean of the prior. Equivalent results were obtained when data from the learning block, the actual angle of centroid of the cloud of dots, was used. However, using only testing data allows for a fair comparison of all models, and allows us to simulate the models even in the absence of behavioral data.

Computing the gradient

To compute the partial derivative of the error function, $e\left(\theta^i,\theta_p,\sigma_p\right)$, we observe that

$$e\left(\theta^i,\theta_p,\sigma_p\right)=\left(\theta^i-\hat{\theta}\left(\theta_p,\sigma_p\right)\right)^2=\left(\theta^i-s_p\theta_f-(1-s_p)\theta_p\right)^2$$

where $s_p=\sigma_p^2/\left(\sigma_p^2+\sigma_f^2\right)$ and $\hat{\theta}$ is defined in **Equation 3.1**.

Applying the chain rule we obtain the partial derivatives:

$$\frac{\partial e}{\partial\sigma_p}=2\left(\theta^i-\theta_p+\frac{\sigma_p^2}{\sigma_p^2+\sigma_f^2}\left(\theta_p-\theta_f\right)\right)\left(\theta_p-\theta_f\right)\frac{\sigma_f^2}{\left(\sigma_p^2+\sigma_f^2\right)^2}2\sigma_p$$

$$\frac{\partial e}{\partial\theta_p}=-2\left(\theta_p-\left(1-s_p\right)\theta_p-s_p\theta_f\right)\left(1-s_p\right)$$

Model fitting.

We fitted the models to the slope and mean data of each subject by minimizing the squared distance to the subjects slope and mean in each direction weighted by the precision (inverse variance, obtained using bootstrapping) of each data point. To account for discrepancies between the learning and testing blocks, both models have an additional scaling parameter that allows us to fit the output of the learning model to subject's prior uncertainty during testing. To compare models (**Figure 3.7**) we bootstrap over the average difference between the weighted RMSE across subjects.

3.4 Results

Here we ask how a noisy perturbation in one, training direction affects reaches into other directions. In particular, we aim to extend movement generalization studies by understanding how both the mean and the variance of a training perturbation affect other movements. Subjects controlled the position of a hidden cursor with their right index finger while their true hand position was occluded by a projector mirror system (**Figure 3.1**). They made reaches from the workspace center to one of eight concentric targets with a visuomotor rotation applied to the hidden cursor position. The visuomotor rotation was drawn randomly each trial from a Gaussian distribution with fixed mean and variance. During *learning* subjects were incentivized (see **Materials and Methods**) to make reaches to one of the targets (training) and received endpoint feedback that allowed them to adapt to the perturbations. During *testing* subjects also made reaches to the other targets, without endpoint error feedback, allowing us to probe generalization. All subjects went through 2 blocks of training, each with a different variance (σ_p: 4^o or 12^o). We measured how the learned variance generalizes, first without perturbing the mean (Experiment 1) and then while also perturbing the mean (Experiment 2).

As subjects adapt to the noisy visuomotor rotations they update their knowledge of both the mean and variance of the perturbations. We can probe subjects' prior uncertainty by providing noisy feedback about the cursor position midway through the movement in the form of a cloud of dots (Körding and Wolpert, 2004). Subjects (n=32 in Experiment 1 and n=8 in Experiment 2) use this midpoint feedback information to correct their movements during each reach (**Figure 3.1**) and they rely more on feedback the more uncertain they are about the cursor position. Computing the *slope* of the negative of final hand position angle (proxy for estimated perturbation) as a function of the perturbation angle (proxy for perturbation sensed via

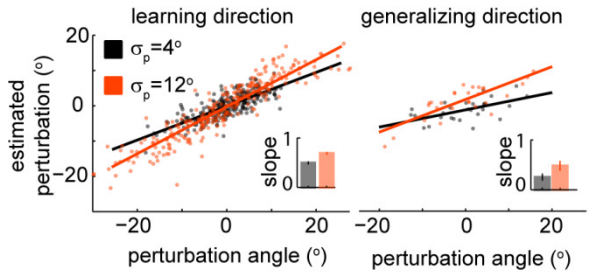
midpoint feedback) provides a measure of the uncertainty that subjects have about the expected perturbations (prior uncertainty) relative to uncertainty about the midpoint feedback (likelihood uncertainty) (Körding and Wolpert, 2004). Intuitively we can see that, if subjects are very certain about the perturbation (low prior uncertainty) then they will tend to ignore the noisy midpoint feedback information and the slope will have a value closer to zero. If on the other hand they have a high prior uncertainty relative to the uncertainty in the midpoint feedback, they will tend to rely only on midpoint feedback and hence the slope will have a value of one. For standard Bayesian integration using Gaussian distributions (Körding and Wolpert, 2004), the slope s_p is given by

$$s_p = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_f^2}$$

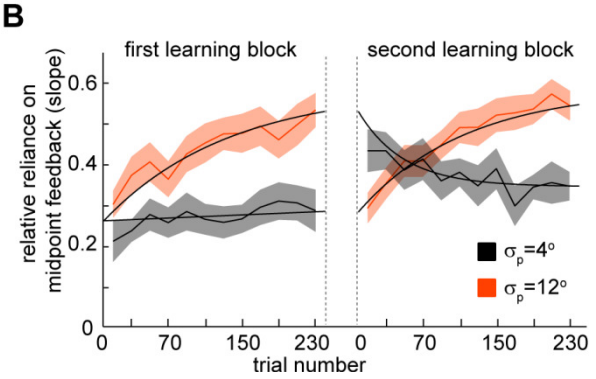
where σ_p^2 and σ_f^2 are the variances of the prior and likelihood distributions, respectively (see Materials and Methods for details). Hence, larger slopes indicate a higher reliance on sensory feedback and higher uncertainty about the perturbations (see **Figure 3.2**A). Whether subjects are Bayesian or not, their slope is a measure of how uncertain they are about the hidden perturbation.

Experiment 1

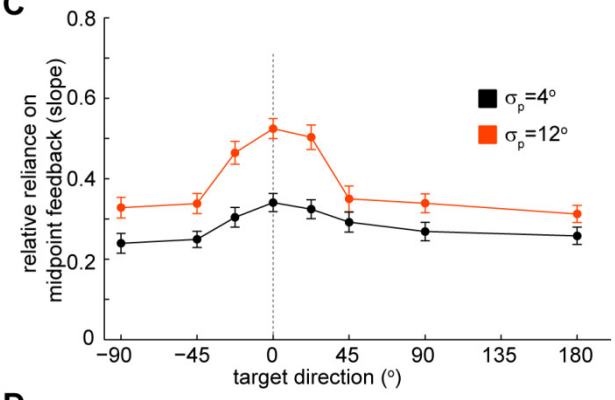
A



B



C



D

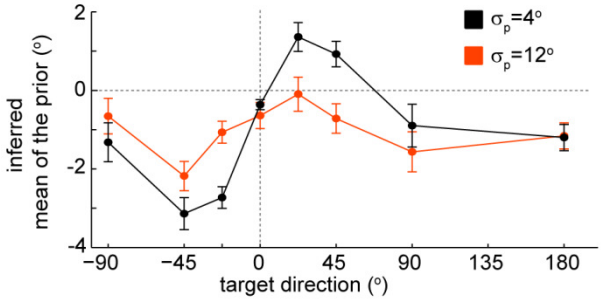


Figure 3.2. Experiment 1: Relative reliance on midpoint feedback (slope) across directions. A, Probing uncertainty in the prior by computing the relative reliance on midpoint feedback. Estimated perturbation as a function of the perturbation angle, for a typical subject, in the learning direction and in one generalizing direction ($+90^{o}$) during the testing phase. Solid lines denote linear fits to the data. *Insets*: Slope or relative reliance on midpoint feedback (±SE) during movements in that direction for this subject. B, Learning for two groups of subjects: subjects that started with the low variance condition and subjects that started with high variance condition. Colored lines are average slopes (±SEM) across subjects considering bins of 20 trials. Black curves are exponential fits. C, Relative reliance on feedback for the two levels of prior uncertainty as a function of target direction relative to learning direction. Mean (±SEM) of slopes across all subjects. D, Inferred mean of prior for the two levels of uncertainty as a function target direction relative to learning direction. Mean (±SEM) of slopes across all subjects.

Experiment 1

We first wanted to know how uncertainty generalizes with a zero mean perturbation. We find that subjects learn about the variance and exhibit smaller slopes for the small variance condition than for the high variance condition (**Figure 3.2**B). This is what we should expect since a smaller slope implies that the subject relies less on the midpoint feedback and, hence, that the subject is more certain a priori about the hidden cursor position. Furthermore, learning curves under the same variance condition converge to the same value during the *learning* phase, no matter which condition subjects started in and appear to saturate before we assess generalization. We do not find a significant difference between the slopes in the two groups of subjects after training ($F_{(1,472)}$=0.95, p=0.33, four-way nested ANOVA over subject, variance, group, target direction where subject is nested in group). By the end of each learning block subjects have adapted to the new

variance condition. In the following analysis we thus combine data across groups to ask how subjects generalize this learned variance.

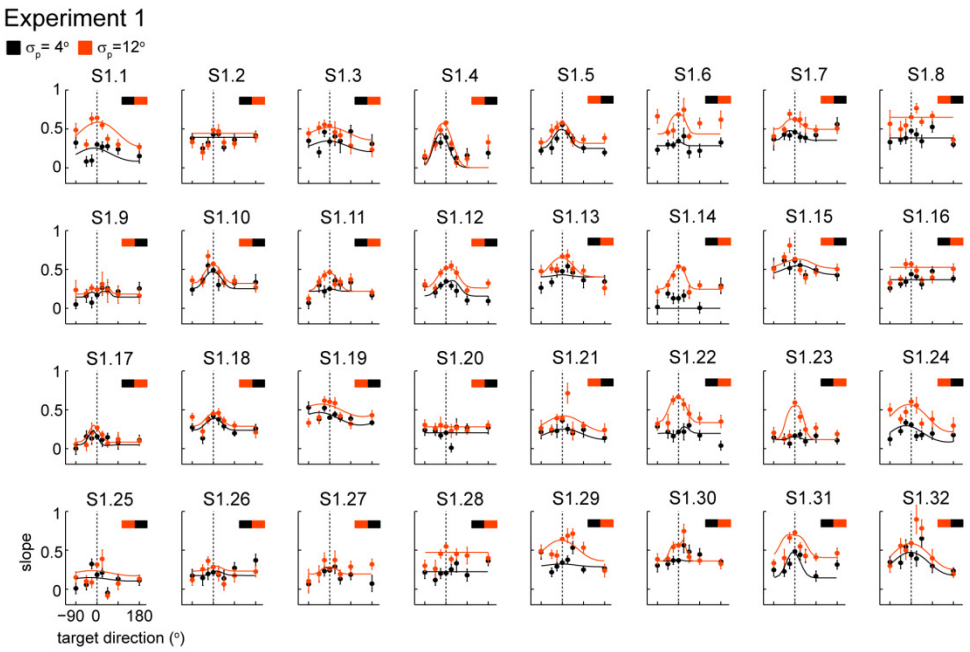


Figure 3.3. Individual subject slope data for Experiment 1. Error bars are ±SEM (bootstrap). Lines are Model 2 (see Models) fits to individual subjects. Black and orange bar (inset) indicates the order in which the different uncertainty blocks were presented to each subject.

To examine generalization of uncertainty we quantify subjects' reliance on midpoint feedback (as measured with the slope) as a function of the direction of movement. We find that the reliance on midpoint feedback (**Figure 3.2C**; see **Figure 3.3** for individual subject data) is significantly different between the two variance conditions ($F_{(1,31)}$=65.13, p<10E-8, two-way repeated measures ANOVA) and slopes for the high variance condition are higher than those for the low variance condition for movements into all directions (p≤0.006, for every direction, paired t-tests, n=32). Uncertainty in

the prior (as measured by the relative reliance on midpoint feedback) increases in all directions and decays with increasing distance from the learning direction. Unlike the mean (Experiment 2 below and Fernandes et al., 2012, but see Taylor and Ivry), uncertainty appears to have a strong global component.

Even though the perturbation had zero mean, we can infer the mean of the subject's prior by analyzing the intercept of a linear regression using data from the testing block (see **Materials and Methods** for details). As with the slopes, we do not find a significant difference between the inferred means in the two groups of subjects ($F_{(1,472)}$~0, p~0.97, four-way nested ANOVA) which allows us to pool the data of both groups. We find an interesting asymmetry consistent with use-dependent learning/adaptation theory (Diedrichsen et al., 2010; Huang et al., 2011; Verstynen and Sabes, 2011); in the reaches towards targets that neighbor the learning target, hand movements are biased towards the learning target and this bias decays with distance from the learning target (**Figure 3.2**). Furthermore the bias is stronger in the low variance condition ($p<0.001$ for both 22.5^{o} target directions, paired t-tests) when movements tend to be less variable and hand position covers a narrower region. We observed signs of a similar effect in a previous study exploring generalization of the mean (Fernandes et al., 2012). We can see it clearly here in the absence of mean adaptation, and where possibly the large number of subjects and the increased complexity of the task (reduced cognitive strategies and across subject variability) makes the effect more observable. These results suggest a weak use-dependent learning effect in this experiment.

Experiment 1 is the analogous for uncertainty of previous generalization studies that measured the generalization of fixed visuomotor perturbations (Krakauer et al., 2000), i.e., the generalization of the mean of the prior with zero imposed variance/uncertainty. In a previous study we showed that the generalization of the mean seems to be unaffected by changes in prior

uncertainty (Fernandes et al., 2012). Fully understanding the generalization of uncertainty requires some understanding of how simultaneously perturbing the mean affects the generalization of uncertainty. Experiment 2 aims to characterize these interactions and differences between generalization of the mean and variance.

Experiment 2

In Experiment 2 our aim is to characterize how the mean of a perturbation affects the generalization of uncertainty. As in Experiment 1, it is important to quantify the effects of the different perturbation variances and to determine whether training order (i.e., high-to-low vs low-to-high variance) matters. Subjects (n=8) readily learned the perturbation variance (**Figure 3.4**A), even with the addition of a non-zero mean perturbation, and we found that, in block 1, the reliance on midpoint feedback (slope) (**Figure 3.4**B) is significantly different between the two variance conditions ($F_{(1,8)}$=19.72, p=0.02, two-way repeated measures ANOVA). However, in Experiment 2 we found a significant difference in reliance on midpoint feedback between the two groups of subjects after training ($F_{(1,112)}$=8.7, p=0.004, four-way nested ANOVA -- subject, variance, group, direction where subject is nested in group). In particular, there are signs of interference in Block 2 (**Figure 3.4**B, C). As savings and interference are a hallmark feature of motor learning (Brashers-Krug et al., 1996; Krakauer et al., 1999) it is not surprising that we should also see them here. Because the order of training now matters we will present and analyze the data from the two blocks separately.

Experiment 2

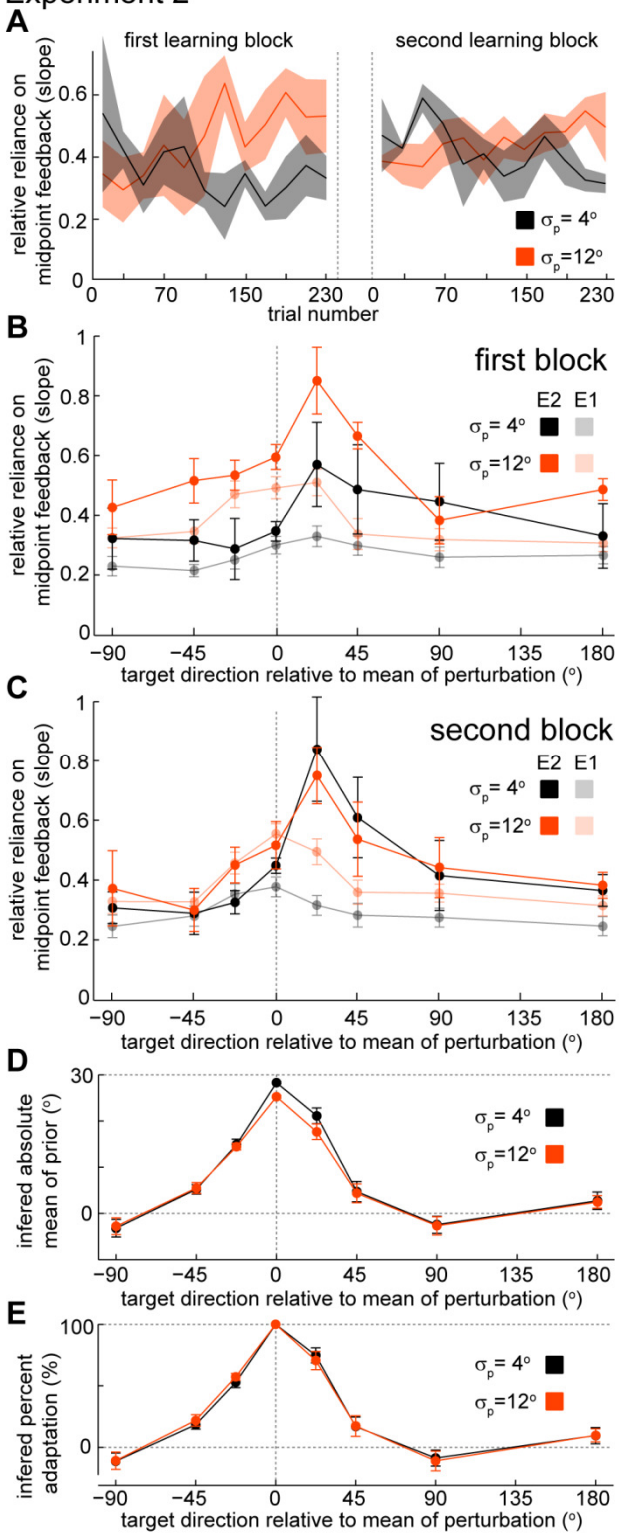


Figure 3.4. Experiment 2. A, Learning of uncertainty for the two groups of subjects: subjects that started with the low variance condition and subjects that started with high variance condition. Colored lines are average slopes (±SEM) across subjects considering bins of 20 trials. B–C, Relative reliance on midpoint feedback for the two levels of prior uncertainty as a function of target direction relative to mean of perturbation (±30^{o}) following the first (B) and second (C) learning blocks. Mean (±SEM) of slopes across all subjects (n=8) in Experiment 2 (opaque solid lines). The transparent lines are the results of Experiment 1 (same data as in Figure 3.2C). D, Inferred mean of prior (±SEM) in the generalizing directions relative to the learning direction (baseline was not measured in this experiment and hence not taken into consideration in this quantification). E, Inferred percent adaptation (±SEM) for the mean in the generalizing directions relative to the learning direction.

In contrast to Experiment 1, here we find a strong asymmetry in the generalization of uncertainty (**Figure 3.4**B, C; see **Figure 3.5**A for individual subject data). The generalized uncertainty as measured by the relative reliance on midpoint feedback is higher than expected for the neighboring targets on one of the sides of the training direction, even higher than the learned uncertainty in the training direction. These directions of higher uncertainty correspond to the opposite direction to where the hand has to move to correct for the perturbation -- that is, the direction of the mean of the perturbation. These are the directions where the midpoint visual feedback is more often displayed during early learning. Furthermore this asymmetry is observed consistently across subjects (**Figure 3.5**A, C) and seems to be robust to any subject-specific cognitive strategies (Taylor and Ivry). We find that when changed simultaneously, the mean and variance of perturbations have asymmetric effects in the generalization of the variance of the prior over those perturbations.

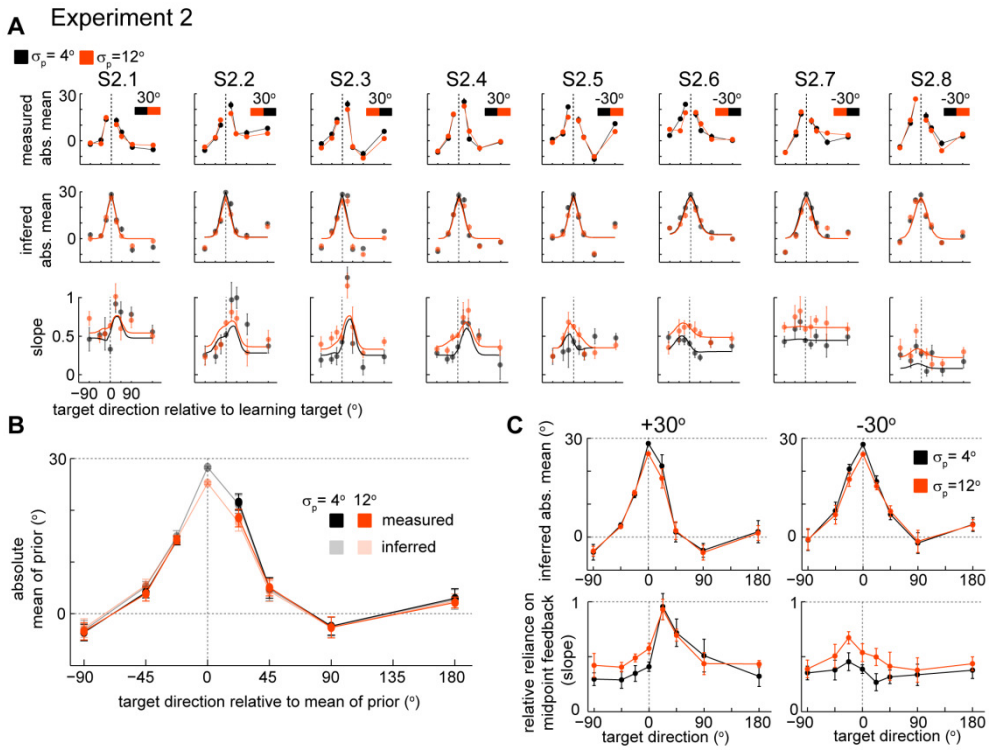


Figure 3.5. Individual subject data for Experiment 2, and further data analysis. A, Error bars are ±SEM (bootstrap). Black and orange bar (first row, inset) indicates the order in which the different uncertainty blocks were presented to each subject. Lines in second and thirst row are Model 2 (see Models) fits to individual subjects. B, Generalization of mean measured during *testing mean* sub-block (opaque), compared with the inferred generalization of the mean (transparent, same as Figure 3.4D during *testing uncertainty* sub-block. C, Generalization of mean and of relative reliance on midpoint feedback (slope) separated by sign of the mean of the perturbation. The asymmetry in generalization of uncertainty (slope) was stronger for the right handed subjects (lower panels).

Since we find a surprising asymmetry in the generalization of variance in Experiment 2, it is reasonable to ask whether manipulating mean and variance simultaneously has a similar effect in the generalization of the mean. We find that the generalization of the mean angular perturbation is

local, with a width of about 30°, similar to what has been reported in previous studies (Fernandes et al., 2012; Krakauer et al., 2000) (**Figure 3.4**D, E and **Figure 3.5**B). As in previous studies, with similar center-out reaching designs (Fernandes et al., 2012), generalization to targets at a ±90° angular distance from the learning target is not significantly different from zero ($p>0.13$ for the ±90° targets in both uncertainty conditions, t-tests). Furthermore, in agreement with Experiment 1, in the directions that neighbor the learning target we observe an asymmetry consistent with use-dependent learning. Note that the use-dependent asymmetry, although reflected as an asymmetric generalization pattern in the mean, can be interpreted as movements close to the training direction being attracted by the direction where training occurred.

Models

If the amount of generalization depends only on similarity between contexts and context is symmetric around target then we would not expect to see an asymmetric pattern in the generalization of variance. In practice, however, the coordinate systems in which subjects try to solve the problem can have an influence on the generalization patterns. To allow for this possibility we hypothesized that the asymmetry could have arisen from a context that is not target centered. Do subjects learn about visual feedback position (Taylor et al., 2012) when generalizing uncertainty?

To see if the data is consistent with this hypothesis and to implement models where feedback position is relevant, we need to consider the distribution of learning data. One natural way of implementing such a model is in terms of online gradient descent. Every trial, one goal of the movement system may be to update certain parameters so that future movements will be better -- we want to go down the gradient of errors (Taylor et al., 2012; Thoroughman and Shadmehr, 2000). We thus implemented two online

learning models that take the perturbations imposed during the learning trials. These models implement gradient descent on the value of, assumed (direction dependent) mean and variance in order to minimize the squared error between target angle and final cursor position angle of each trial (see Materials and Methods for details). Model 2 uses a coordinate system for generalization of variance related to the position of visual feedback while Model 1 uses only target-centered coordinates. This way of phrasing the problem allows us to consider the effect of the candidate coordinate systems on learning and generalization.

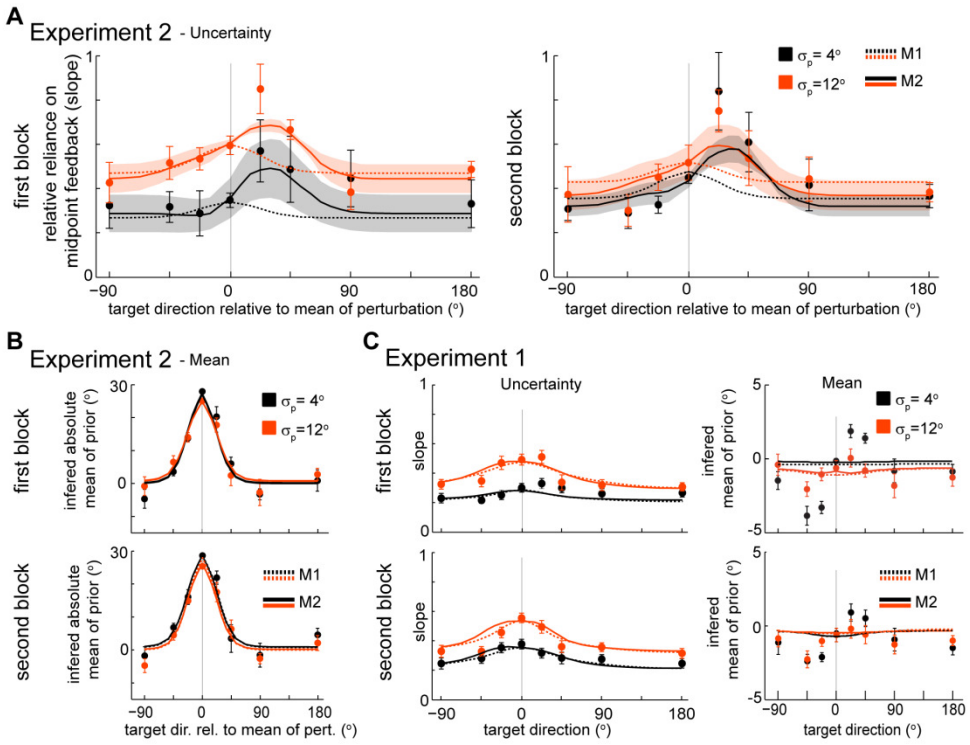


Figure 3.6 Online learning models for different reference frames. A, Models fit (shaded area is ±SEM for M2 fits) for the slope data of Experiment 2 (same as Figure 3.4B, C, opaque). Error bars are ±SEM. B, Models fit for the mean data of Experiment 2. C, Models fit for mean and slope data (same as Figure 3.4B, C,

transparent) of Experiment 1. Error bars are ±1SEM. Lines are average across subjects of individual fits. Error bars are ±1SEM.

We find that Model 2 captures the generalization patterns of both experiments (**Figure 3.6**; see **Figure 3.3** and 3.5A second and third rows for individual subject fits). Importantly, Model 2 was able to capture the asymmetric generalization of uncertainty of Experiment 2 (**Figure 3.6**A) and, simultaneously, explain the data in Experiment 1 (**Figure 3.6**C) – except for the use-dependent effect. We find that, while none of the models is significantly better for Experiment 1 ($p>0.14$ for uncertainty and for mean; bootstrap of RMSE of individual fittings across subjects, see **Figure 3.7**A, C), Model 2 is better than Model 1 for Experiment 2 ($p<10^{-4}$ for uncertainty; bootstrap individual fittings across subjects, see **Figure 3.7**B, D). Although lacking a normative interpretation/justification, using different reference frames for mean and variance and using gradient descent learning accurately captures the generalization patterns across experiments.

3.5 Discussion

Here we examined how priors over a stochastic visuomotor perturbation generalize. We examined in particular, how prior uncertainty, that is, knowledge of the trial-by-trial *variability*, generalizes. We first tested generalization when we changed only the variance of the distribution of rotation perturbations and not the mean. We found that, similarly to standard generalization of visuomotor rotations, generalization of uncertainty has a local component. However, unlike the mean, it affects movements into all directions. We then tested how uncertainty generalizes when we introduce a stochastic perturbation with non-zero mean. We observed asymmetric

generalization that is qualitatively consistent with a descriptive, online learning model that assumes that mean and variance generalize according to different reference frames.

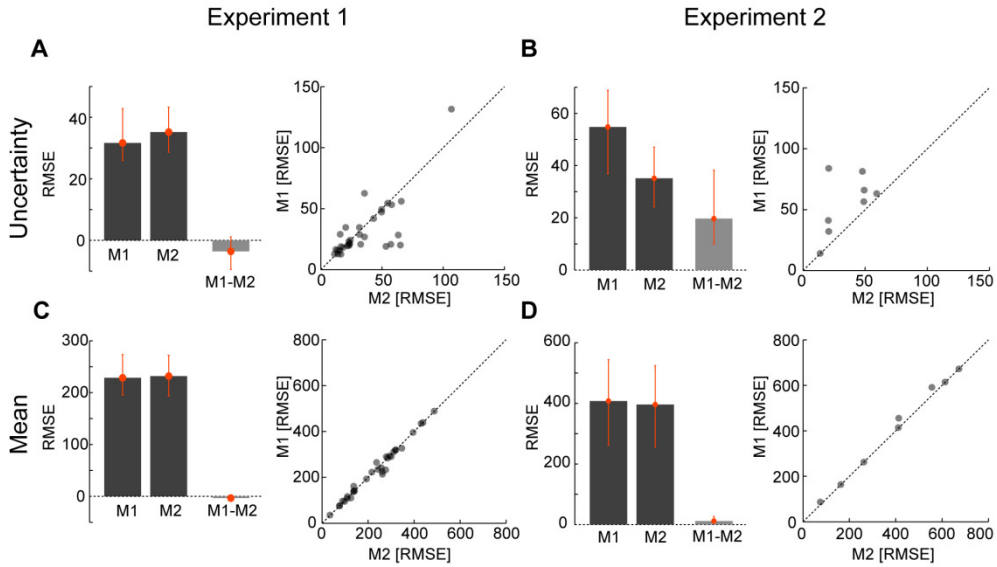


Figure 3.7. Model comparison. A–D, Model comparison for Models 1 and 2 for Uncertainty (A and B) and for Mean (C and D). (Left) Weighted root mean square error (RMSE) across subjects (95% confidence intervals, bootstrap) of each model and of the difference between models for each subject. (Right) scatterplot of the RMSE for each subject for Models 1 and 2.

In movement research, generalization experiments are usually interpreted as being directly related to neuronal tuning properties (Krakauer et al., 2000a; Thoroughman and Shadmehr, 2000) (but see Pearson et al., 2010; Taylor and Ivry). Under this interpretation they constrain our conceptualization of neural computation and reveal a great deal about the neural basis of sensorimotor integration. We had seen evidence for some

independence in representation of mean and variance of priors in previous studies when we showed that uncertainty does not affect the width of generalization of the mean (Fernandes et al., 2012). The results of this study indicate that knowledge of the variance of external perturbations might be represented in a way that is distinct from the knowledge about the mean – both the extent of generalization and reference frames appear to differ.

The degree to which the brain is "Bayesian" has been extensively debated over the last decade (Doya, 2007). Many studies have shown that the brain achieves Bayes-like behavior for familiar tasks (such as reaching) and that this behavior stems from ongoing learning (Berniker et al., 2010). Such general-purpose Bayesian behavior may result from a variety of non-Bayesian/heuristic neural representations. Alternatively, Bayesian ideas may be far more fundamental to the organization of the brain in the sense that there is something Bayesian about the neural code itself. For example, spikes in populations of neurons might directly represent probability distributions, including their means and variances (Deneve, 2008; Fiser et al., 2010a; Hinton and Sejnowski, 1983a; Hoyer and Hyvärinen, 2003; Ma et al., 2006; Ma, 2010; Sahani and Dayan, 2003; Soltani and Wang, 2009; Wu et al., 2003; Zemel et al., 1998). None of these "Bayesian Brain" theories explicitly predicts generalization of uncertainty and, generalization is probably related to underlying neural representations in a more complex way than generally assumed in motor control research. However, dissociation between generalization of mean and variance emerges immediately from our results and produces an important challenge to extensions of "Bayesian brain" theories to generalization.

The lack of computational predictions for generalization of priors, and of uncertainty in particular, is mirrored in experimental work where the focus both in behavioral as well as in electrophysiological studies in motor control has been on the generalization and representation of fixed perturbations without any trial-by-trial variability (but see Fernandes et al., 2012 and

Verstynen and Sabes, 2011). Previous work had not indicated that means and variances could generalize differently.

Previous studies have shown that the reference frames for generalization depend on context, and that we expect different generalization patterns if different contexts are imposed (Berniker and Kording, 2008; Brayanov et al., 2012; Taylor and Ivry). Studies that focus on the adaptation of the mean suggested that feedback plays an important role in adaptation (Huang et al., 2011) and generalization (Taylor et al., 2012). Differences in visual error information lead to changes in generalization that can be explained by a neural network that assumes error feedback processing on a set of homogeneous and invariant tuning functions (Taylor et al., 2012). The use of the reference frame of visual feedback for the learning of the mean has thus been shown previously. This study suggests the use of a visual reference frame for the generalization of uncertainty.

The model presented here makes several predictions for future experiments as well. Some studies look at what happens when perturbations are introduced gradually (Berniker and Kording, 2011; Kagerer et al., 1997; Turnham et al., 2012). If Model 2 is correct then we expect gradually introduced perturbation to produce a less asymmetric generalization curve for uncertainty, since the feedback would generally not appear so far away from target direction. Another interesting follow-up experiment would be to do the same set of experiments for visuomotor gain instead of visuomotor rotation - whose mean has been shown to generalize globally in minimal uncertainty conditions (Krakauer et al., 2000).

The fact that the task was more complex than previous studies (Fernandes et al., 2012) allowed us to infer the mean of the prior with smaller variability across subjects. We find clear signs that movements are biased towards typical directions of previous hand movements, which is consistent with the use-dependent learning/adaptation hypothesis (Diedrichsen et al., 2010;

Huang et al., 2011; Verstynen and Sabes, 2011). We find this in both experiments and it is particularly evident in Experiment 1 where the mean of the distribution of stochastic perturbations was zero; this use-dependent asymmetry scales with the uncertainty level and exists even when there is zero mean perturbation. Although our model captures features of generalization patterns for both mean and uncertainty, it does not capture this use-dependent aspect of the generalization. Future work could account for these effects by incorporating a hand centered reference frame or other "model-free" learning processes (Huang et al., 2011).

In fact, even though Model 2 is consistent with observed symmetry we cannot exclude that it might be caused by other mechanisms. It is not an unreasonable hypothesis that the same mechanism responsible for the use-dependent asymmetry in the generalization of the mean is responsible for the asymmetry in the generalization of uncertainty. If this is true it happens in a way that is not obvious to us and future research could try to address it.

Where priors come from and how they are represented are fundamental questions in learning and behavior. As we never experience the same situation twice, constructing priors depends crucially on our ability to generalize. However, generalization in both perception and action is a result of how the brain represents the external world. In perception research, studies that hypothesize priors based on the statistics of natural scenes (Burge et al., 2010; DiMattina et al., 2012; Geisler et al., 2001; Roth and Black, 2005) generally assume certain invariances where global generalization occurs along many dimensions of the stimulus. When calculating orientation priors, for instance, color and contrast are assumed to be irrelevant and only the statistics over orientation are considered important (Girshick et al., 2011). In movement research, it is generally assumed that the system is invariant to the content of the visual scene and that generalization only depends on (angular) distance (Krakauer et al., 2000), velocity (Goodbody and Wolpert, 1998) or the way an object is held

(Ingram et al., 2010) (but see Taylor et al., 2012). For both perception and action, the nature of the underlying representations determines the shape of generalization. Quantifying the generalization of priors, taking uncertainty into account, allows new ways of understanding these representations.

4. Prior Beliefs and Decision-Making Under Uncertainty

Daniel Acuna*, Max Berniker*, Hugo L Fernandes*, and Konrad P Kording

* Equal contribution

Article in preparation

Author Contributions. Conceived and designed the experiments: DA MB HLF and KPK. Performed the experiments: MB, DA and HLF. Analyzed the data: DA MB HLF and KPK, Implemented the analysis: DA with the help of MB and HLF. Implemented the experiment: DA and MB with the help of HLF. Wrote: DA MB HLF

4.1 Summary

The two-alternative-forced-choice (2AFC) paradigm and the resulting just-noticeable difference (JND), are generally assumed to quantify sensory uncertainty independent of a subject's beliefs (i.e. their prior). This interpretation is consistent with the maximum a posteriori (MAP) decision theory, according to which subjects choose the option most probable (using a posterior distribution to represent subjective belief). However, a host of alternative decision-making theories, including sampling and matching, predict choices should be influenced by prior beliefs. Here we mathematically examine the predictions of these different theories and, using the results from an interleaved estimation and 2AFC task, find no influence of the subjects' prior beliefs on their measured JNDs. These results are consistent with the MAP hypothesis, arguing against sampling theories of decision making. We propose that the 2AFC task is not a

straightforward tool for measuring subjects' sensory precision, but rather a probe for theories of the neural representation of uncertainty.

4.2 Introduction

Every decision we make is a choice, constrained by our options and based on limited and uncertain information. For example, which lane on the highway should you drive in? which line at the grocery store should you wait in? Sometimes our options have clear differences, while other times the options and their outcomes are nearly identical and hard to distinguish. Though these decisions are commonplace and indicative of how we make choices in general, how people choose between multiple uncertain options remains largely unknown.

There are multiple prominent theories describing how people make decisions under uncertainty. Normative theories assume that sensory information (generating a likelihood function) is optimally combined with our expectations (a prior distribution) into a belief (posterior distribution) over the outcome of our choices (e.g. the probability getting home early for each lane we drive in). Various theories on decision-making differ only in how a choice is made from this posterior probability. One alternative is that people choose the most probable alternative (the maximum a posteriori probability---MAP hypothesis). Alternatively, people may not be able to compute the most probable outcome, and must approximate it instead. Under this hypothesis, subjects draw one or more sample choices from their posterior, and compute the best sample statistic, the so-called "sampling hypothesis" (Vul et al., 2009). If subjects use only one sample, they are said to be "matching" (Vulkan, 2000; Wozny et al., 2010). Though both MAP and the sampling hypotheses offer very different predictions, they can both be viewed as optimal under the right assumptions (e.g., for MAP (Duda et al., 2012); for

sampling (Sakai and Fukai, 2008; Vul et al., 2009; Wozny et al., 2010)) and thus equally normative in their description of decision-making.

The two-alternative forced choice (2AFC) task has been the workhorse of psychophysics and decision-making experiments for the last 150 years (Green and Swets, 1966). In this task, subjects are presented with two alternatives and forced to choose between them based on some experimentally defined attribute. For example, subjects may be asked to decide which of two flashes of light displayed on a screen is further to the left. By controlling the discrepancy between these cues (location of light flashes), experimenters can obtain a psychometric curve: the probability of a subject's response given the discrepancy between cues. This curve can then be used to quantify the just-noticeable-difference (JND), which is related to how different must the two cues be before subjects can reliably tell them apart (Green and Swets, 1966). Due to its simplicity and experimental benefits, the 2AFC task is used in a broad variety of sensory and cognitive domains to measure the JND.

In the majority of circumstances, the 2AFC task and psychometric curve are assumed to characterize sensory precision (Ernst and Banks, 2002; Fetsch et al., 2011; Girshick et al., 2011; Stocker and Simoncelli, 2006; Tassinari et al., 2006). Importantly though, the resulting JND is thought to be independent of an individual's prior beliefs. This is beneficial for several reasons, not least of which because it precludes the possibility that the measured JND can be influenced by subject biases or experimental circumstances. However, how people perform the 2AFC task, and what the JND measures, relies crucially on how people make decisions under uncertainty; depending on the subjects' strategy, the JND may or may not be influenced by a subject's prior, confounding its interpretation as a measure of sensory precision. This confound raises concerns for the great number of studies that rely on psychometric curves and measured JNDs obtained in the 2AFC task.

Here we mathematically examine the MAP and sampling theories of decision-making and their influence on the 2AFC task and resulting JND. We demonstrate how if, after combining sensory and prior information, subjects choose the option that maximizes their posterior (i.e. the MAP answer) then the JND correctly measures a subject's sensory precision. However, if subjects choose according to the sampling hypothesis, the psychometric curve and JND measure something altogether different that depends on their prior. We then exploit this result to design an experimental paradigm to test how people make uncertain decisions. Using an interleaved estimation and 2AFC task, we measure subjects' prior beliefs and their JNDs. In our task we found that changes in a subject's prior had no measurable influence on their JND, consistent with MAP decision-making. Our results thus support the traditional interpretation of the psychometric curve as measuring sensory precision, independently of prior knowledge. However, we propose that in general, the 2AFC task is not a straightforward tool for measuring a subject's sensory precision. Instead, the 2AFC task can be used to probe and falsify theories of decision-making under uncertainty (Gold and Ding, 2013; Liston and Stone, 2008; Palmer et al., 2000).

4.3 Results

Our aim was twofold. First, we sought to examine several prominent theories for decision-making under uncertainty, and in particular how they would influence a psychometric curve. Second, we sought to design an experiment capable of testing these predictions. These analyses would clarify the implicit assumptions behind the common use of the 2AFC task, and the experiment would test whether the JND is in fact a measure of sensory precision, or merely a phenomenon of dubious distinction. Below we briefly introduce the relevant decision variables, the conventional 2AFC interpretation and determination of the JND. This will provide the

groundwork for the subsequent mathematical analysis of the 2AFC task under the sampling and maximizing theories. We then describe the experiment and present the experimental results and their analyses.

Decision-making Theory and the 2AFC

In the 2AFC task, people are asked to make a binary choice based on two experimentally imposed cues, which we refer to as c_1 and c_2. They could be two differentially illuminated flashes of light or two sounds of different pitches. For instance, suppose that the task is to decide whether the second cue is greater in value than the first cue (i.e. is $c_2 > c_1$ true of false?) Since the answer to this or any 2AFC question is binary, we can describe the answer with the random Bernoulli variable, $z \in \{0,1\}$. By asking subjects to perform many trials of this task, while systematically manipulating the difference between c_2 and c_1, we obtain a psychometric curve. This curve characterizes the probability that a subject chooses one cue over the other as a function of the difference between them, δ. One way of describing this relationship is with the cumulative normal distribution,

$$P(z=1 \mid \delta) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\delta - \mathrm{PSE}}{\sqrt{2}\sigma_{\mathrm{JND}}}\right)\right] \qquad 4.1$$

where erf denotes the error function. The standard deviation, σ_{JND}, often referred to as the just-noticeable difference (JND), describes the behavioral precision in discriminating two cues. The experimental and behavioral aspects of the 2AFC task are well-studied and analyzed and can be found elsewhere (Green and Swets, 1966).

The psychometric curve and JND presented above quantify subject behavior, but do not describe how choices are made. This decision process is formalized in essentially the same manner by many normative theories. Due to the noise imposed by the experimental settings as well as to the inherent noise in the human sensory apparatus, the two cues give rise to uncertain sensations, which we shall label as s_1 and s_2. Formally, our sensory information induces the likelihood of every possible value of the cue for each sensation: $P(s_1 \mid c_1)$ and $P(s_2 \mid c_2)$. Our percepts can be thought of as another probability distribution, $P(c \mid s)$, which is our posterior belief of each cue probability given our sensation of them. Applying Bayes' formula, we find,

$$P(c \mid s) \propto P(s \mid c)P(c)$$

where the prior, $P(c)$, is our subjective expectation based on a lifetime of experiences. Hence, our perception of the cue is a function of both our senses, and our prior belief in what we expect the cue to be. Now we can formally interpret the 2AFC task as a decision, z, based on two probability distributions, $P(c_1 \mid s_1)$ and $P(c_2 \mid s_2)$. This allows us to predict different distributions, $P(z \mid \delta)$, for different candidate decision-making theories and determine when, if ever, the JND is influenced by prior beliefs. Additionally, by comparing these predictions against subject behavior, we can use the 2AFC task as a tool to corroborate or falsify decision-making theories.

Maximum a posteriori (MAP) decision-making

Under the MAP hypothesis, subjects make their decision based on the most probable choice; that is, in choosing which cue is larger, one simply

compares the most probable value of c_2, with the most probable value of c_1. Mathematically, the choice is defined as follows:

$$z = \begin{cases} 1 & \text{if } \arg\max_{c_2}[P(c_2 \mid s_2)] \geq \arg\max_{c_1}[P(c_1 \mid s_1)] \\ 0 & \text{otherwise} \end{cases}$$

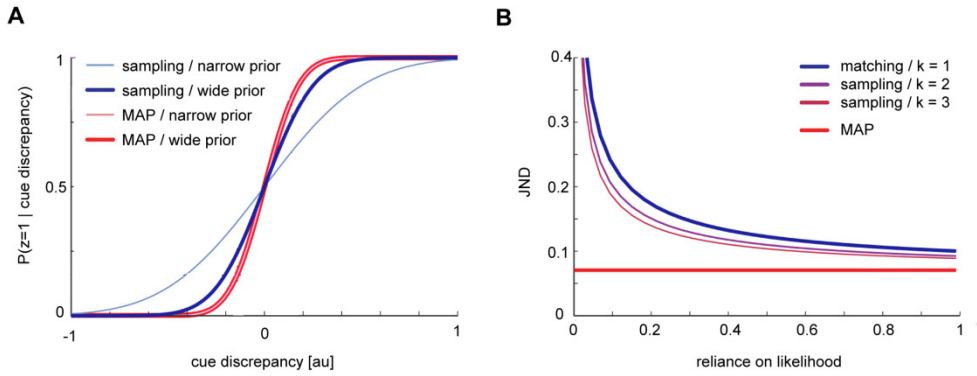


Figure 4.1. Predictions of the different decision making theories. (A) Psychometric curves predicted for the sampling models and the MAP model under different prior conditions. Note that the MAP/narrow prior and the MAP/wide prior lines overlap. (B) JND of the psychometric curves as a function of relative reliance on likelihood (proxy for prior uncertainty). The predictions for the sampling models (with different number of samples, k) and the MAP model are shown.

By assuming functional forms for the likelihood $P(s \mid c)$ and prior $P(c)$, we can then obtain $P(z \mid \delta)$, by integrating over our senses, s. In particular, if we assume normal distributions (see Supplemental Information) we get,

$$P(z = 1 \mid \delta) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\delta}{2\sigma}\right)\right],$$

where σ is the standard deviation of the likelihood. To be clear, if subjects choose according to the MAP hypothesis, the function above accurately models their behavior, whereas Equation 4.1 is used to fit their behavior. Note that the resulting psychometric curve (and JND) are independent of the subject's prior (see also **Figure 4.1**A). Furthermore, by comparing terms with Equation 4.1 we can define the experimentally derived JND in terms of the precision of a subject's likelihood:

$$\sigma_{\text{JND}}^{\text{MAP}} = \sqrt{2}\sigma$$

The assumptions we made to derive these results are the implicit assumptions that most studies make when the experimental JND is interpreted as a measure of the subject's sensory accuracy. However, as we demonstrate below, alternative decision-making theories predict distinct results.

Decision-making theory	JND
(1) MAP	$\sigma_{\text{JND}}^{\text{MAP}} = \sqrt{2}\sigma$
(2) sampling	$\sigma_{\text{JND}}^{\text{sampling}} = \left[\sqrt{2}\sigma\sqrt{(k+1)\sigma_c^2 + \sigma^2} \right] / \sqrt{k}\sigma_c$
(2) matching (k=1)	$\sigma_{\text{JND}}^{\text{matching}} = \sqrt{2}\sigma\sqrt{2 + \sigma^2 / \sigma_c^2}$
(3) sampling with $\sigma_c \to +\infty$	$\sigma_{\text{JND}} \underset{\sigma_c \to +\infty}{\to} \sqrt{\frac{k+1}{k}}\sqrt{2}\sigma_c$

Table 1. Different predictions for the measured JND, according to MAP and sampling/matching decision-making theories.

Sampling-based decision-making

In contrast with the above result, the sampling-based decision-making hypothesis proposes that choices are based on an approximation to the most probable outcome. This approximation is computed by “sampling” from

the two posterior distributions, and comparing the averages. Mathematically, we can express the choice as follows:

$$z = \begin{cases} 1 & \text{if } \bar{c}_2^k \geq \bar{c}_1^k \\ 0 & \text{otherwise} \end{cases}, \qquad 4.2$$

where $\bar{c}_1^k$ and $\bar{c}_2^k$ are the sample means computed by drawing k samples from the posterior distributions, $P(c_1 \mid s_1)$ and $P(c_2 \mid s_2)$. Again, by assuming distributions for the likelihood and prior we can then obtain $P(z \mid \delta)$. If we assume normal distributions for the likelihood and the prior, $P(c_1) = P(c_2) = N(\mu, \sigma_c^2)$, we find,

$$P(z = 1 \mid \delta) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\sqrt{k}\sigma_c \delta}{2\sigma\sqrt{\left((k+1)\sigma_c^2 + \sigma^2\right)}}\right)\right]$$

We note several features of this result, first of which is the appearance of terms from the prior (see Figure 4.1A). Again by matching terms with Equation 4.1, under this hypothesis the experimentally derived JND is not merely a subject's sensory accuracy, but rather a combination of both sensory and prior uncertainties:

$$\sigma_{\text{JND}}^{\text{sampling}} = \frac{\sqrt{2}\sigma\sqrt{(k+1)\sigma_c^2 + \sigma^2}}{\sqrt{k}\sigma_c} \qquad 4.3$$

This result is in stark contrast with the traditional interpretation of the 2AFC task. We see that when a subject's prior is certain (relatively small σ_c), the JND increases (see Figure 4.1B). Intuitively, we interpret this as follows: as the prior becomes more and more certain, sensory information becomes less relevant and the posterior belief is more closely aligned with the prior.

Therefore, distinguishing a difference between the two cues requires increasingly large differences.

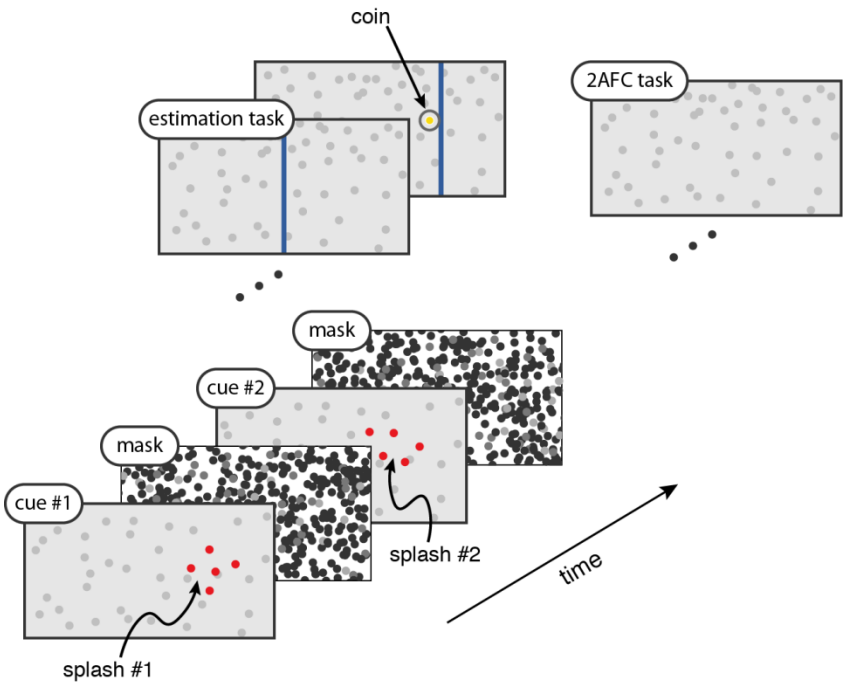


Figure 4.2. Experimental protocol. Subjects were shown two sets of "splashes" (5 dots) and then randomly presented with one of two tasks; either (A) the estimation/coin-catching task or (B) the 2AFC task. (A) On the estimation trials subjects were prompted to place a net (vertical blue bar) where they estimated the hidden coin position to be. (B) On the 2AFC trials subjects had to estimate which of the two hidden coins landed more to the right.

In the limit of an infinite number of samples, the JND under the sampling hypothesis is equivalent to the MAP prediction. In the limit of an infinite variance prior, the JND tends to

$$\sigma_{\mathrm{JND}} \underset{\sigma_c \to +\infty}{\to} \sqrt{\frac{k+1}{k}}\sqrt{2}\sigma_c$$

which is different from the MAP prediction (if k is finite). We also note the special case where $k=1$ is the so-called matching hypothesis (Vulkan, 2000; Wozny et al., 2010). This scenario is equivalent to the hypothesis that subjects choose between their choices with a rate that is proportional to the probability of being correct; that is, $z=1$ with probability $P(c_2 > c_1 \mid s_2, s_1)$. By observation the JND is now:

$$\sigma_{\text{JND}}^{\text{matching}} = \sqrt{2}\sigma\sqrt{2+\sigma^2/\sigma_c^2}\ .$$

These systematic differences between the MAP and sampling predictions (**Table 1**) suggest a way of using subjects' performance during the 2AFC task to investigate how they make decisions under uncertainty. Below we present the details and results of an experiment design that aims at doing so.

Measuring subjective beliefs

Based on the derivations above, we designed an experiment to manipulate subjects' uncertainty in the prior while simultaneously quantifying changes in their JND. Each of the seven subjects performed the experiment on five separate days. On each day, they performed 2,000 trials, randomly switching between estimation trials (1000/day), where they had to estimate the location of a hidden coin, and 2AFC trials (1000/day), where they had to decide which of two coins was further to the right (see below, and Materials and Methods). Halfway through each day's experiment, the variance of the prior distribution would switch, from large to small or from small to large. These conditions allowed us to test whether subjects' JND changed when the variance of the prior changed.

During estimation trials, subjects' guesses for the location of the hidden coins were used to measure their subjective beliefs, commonly referred to as their prior. By recording how their guesses varied as the location of the evidence (splashes/likelihood) varied, we measure two features: the mean of their prior (see Figure 3A and Materials and Methods) and their reliance on the likelihood relative to the prior ("reliance on likelihood", for short; see Figure 3B, C and Materials and Methods). The reliance on likelihood is an indirect measurement of the variance of a subject's prior (Berniker et al., 2010a; Körding and Wolpert, 2004b; Vilares et al., 2012). With these measurements we could determine if subjects learned the experimentally manipulated distribution of coins.

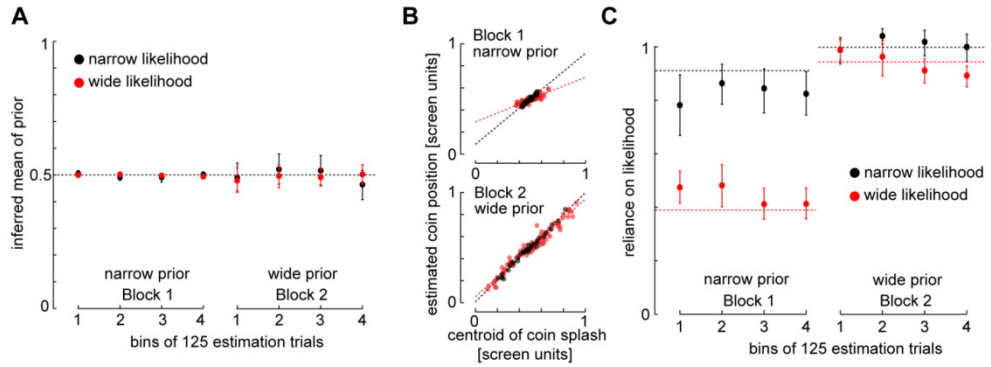


Figure 4.3. Measuring subject's priors. A) Estimated mean of the prior for a typical subject on both blocks of the first day. Error bars are 95% CI (bootstrap). B) Data from the estimation trials for a typical subject on their first day. The slope of a linear regression determines the reliance of the likelihood, a proxy for subject's prior uncertainty. C) The reliance on the likelihood is binned to visualize learning. As subjects learn the prior (dashed lines are theoretical values that correspond to the experimental variances of prior and likelihood). Error bars are 95% CI (bootstrap).

First we wanted to know if subjects learn a prior and whether they take prior and likelihood variances/uncertainties into account when deciding where to place the net. That is, (a) whether subjects learn a different prior in the low prior uncertainty block relatively to the high prior uncertainty block and (b) whether they also take the different uncertainty levels in the likelihood into account. As this estimation paradigm has been used with successful results in previous research (Berniker et al., 2010a; Vilares et al., 2012) we expect this to happen. Indeed, the fitted slope to the data of the estimation trials during the first 250 trials of the first day were significantly different across priors ($F(1,19)=44.2$, $p<0.01$, ANOVA), and likelihood ($F(1,19)=12.8$, $p<0.01$, ANOVA), but not subjects, ($F(1,19)=1.5$, $p=0.22$, ANOVA), suggesting that the experimentally manipulated priors and likelihoods had an effect in subjects behavior. The same result holds for the estimation trials within the last 250 trials of the first day. Subjects typically learned the task already in the first day as is evident from the different behavior for the different conditions. The additional four days were necessary to more precisely quantify their psychometric curves.

Next we wanted to know if subjects learned similar priors across days. This is important if we could combine the data across days. The fitted slopes during the estimation trials during the last 250 trials of each day was significantly different across priors, $F(1,127)=325.9$, $p<0.01$, likelihoods, $F(1,127)=90.67$, $p<0.01$, subjects $F(6,127)=6.44$, $p<0.01$, but not significantly different across days, $F(4,127)=1.26$, $p=0.28$, suggesting that subjects learned similar priors across days. The slopes for the different prior and likelihood conditions did not change significantly across days for each subject.

The average reliance on the likelihood across subjects and days was 0.72 (SE 0.01) for narrow prior and narrow likelihood, 0.43 (SE 0.01) for narrow prior and wide likelihood, 0.95 (SE 0.007) for wide prior and narrow likelihood, and 0.91 (SE 0.006) for wide prior and wide likelihood. These

numbers are significantly different from the optimal values but show a trend qualitatively consistent with the optimal values: 0.91 for narrow prior and narrow likelihood, 0.39 for narrow prior and wide likelihood, 0.99 for wide prior and narrow likelihood, and 0.94 for the wide prior and wide likelihood conditions. The fact that subjects don't learn the exact experimentally imposed prior is not a problem given that what we need is for subjects to learn a sufficiently different prior across condition that allows distinguishing between decision-making theories.

Based on the above findings we conclude that subjects take into consideration the prior and the likelihood uncertainty when making a decision in the estimation task. Importantly they learn a different prior for each imposed prior condition and we are able to obtain a relative measure of each subject's subjective prior. This allows us to examine if changes in their priors influenced their JND's. The results of this examination would provide evidence for either a MAP or sampling decision-making process.

Measuring psychometric functions

As described above, on each day subjects performed 1000 2AFC trials. The data from these trials were used to fit psychometric functions. These subject-specific curves quantify how large the discrepancy between two coin splashes needs to be before subjects can reliably perceive them as distinct. In roughly half of the trials, the two coin splashes had approximately the same size. We denote these trials as *same-likelihood 2AFC trials*. Using this data we could fit a psychometric curve to each subject's responses, and measure their JNDs and PSEs (see Methods). The psychometric curves (using the same-likelihood 2AFC trials) would give us valuable estimates of subject-specific JNDs. These estimates are necessary to test the decision-making theories.

In the remaining 2AFC trials, the two coin splashes had different sizes; *different-likelihood 2AFC trials*. When two different likelihoods are used to make a choice, the probability of a response is a function of both cue locations (not merely the difference between them as it happens in the same-likelihood condition, see Figure 4A); the psychometric function is now a surface (see **Figure 4.4**B and Supp. Information). We used these surfaces to measure how the PSE's changed with cue locations. These surfaces, computed using the different-likelihood 2AFC trials, were used as a valuable control; since the use of a prior makes predictable changes in the way the cues' position affect the shape of the surface (see below and Supp. Information) we can use this data to test whether the prior learned in the estimation trials is used in the 2AFC trails.

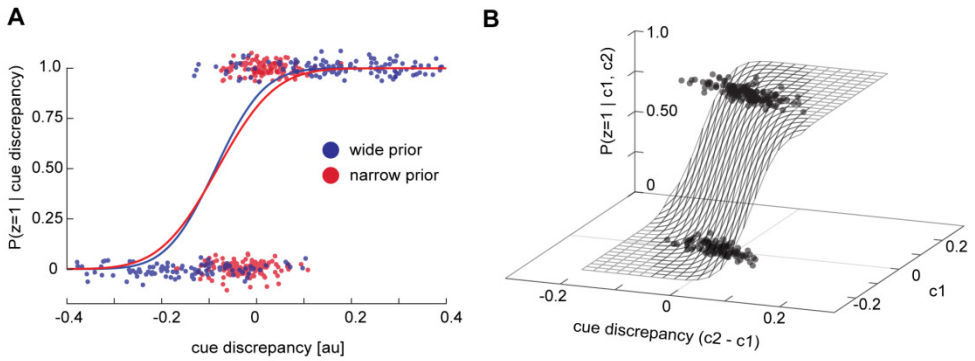


Figure 4.4. Subject-specific psychometric curves. A) Data from the 2AFC trials is used to fit a psychometric curve. Red and blue dots are subject responses during the narrow and wide conditions, respectively. B) When the standard deviations of the two cues are different, the probability of a response being true is a function of the location of the both reference cues and the fit is now a surface.

Subjects use the learned prior during the 2AFC task

A basic assumption of the 2AFC analysis is that subjects use their prior when making choices. However, it could be the case that subjects only use one prior in the estimation trials, and use a different strategy for the 2AFC trials. For example, subjects might simply use the respective centroids to choose between splashes (i.e. a maximum-likelihood estimate), or a different prior and neglect the prior learned during the estimation times all together. To exclude this possibility, we examined the psychometric surfaces obtained in the different-likelihood 2AFC trials. If subjects were using the prior learned during the estimation trials, these surfaces would change in a predictable way.

In the different-likelihoods 2AFC trials, the PSE for both MAP and sampling hypotheses is a function of the cue positions (see Materials and Methods and Supp. Information). In particular, the PSE should change linearly with the cue positions, We can express this in terms of either cue's position, for instance, c_1, and their respective variances:

$$\mathrm{PSE}(c_1) = \left(\frac{\sigma_2^2 - \sigma_1^2}{\sigma_1^2 + \sigma_c^2} \right) c_1$$

where σ_1, σ_2 are the likelihood variances and σ_c is the prior variance. We denote the term $(\sigma_2^2 - \sigma_1^2)/(\sigma_1^2 + \sigma_c^2)$ by *slope of PSE*. Importantly, the prediction is that the slope's magnitude (absolute value) will decrease as the prior's variance increases, allowing us to predict how it should change across the small and large experimental prior variance conditions. We chose c_1 to be the splash with large variance and hence the slope of PSE is predicted to be negative.

In agreement with the predictions, we found that the PSE slope for the small prior was significantly smaller (than for the large priors ($p<0.05$, paired t-test; see **Figure 4.5**A). This result supports the assumption that subjects used

their subjective prior learned during the estimation trials when making their choices during the 2AFC task. We can now finally check if changes in prior affects the JND measured using data from the same-likelihood 2AFC trials.

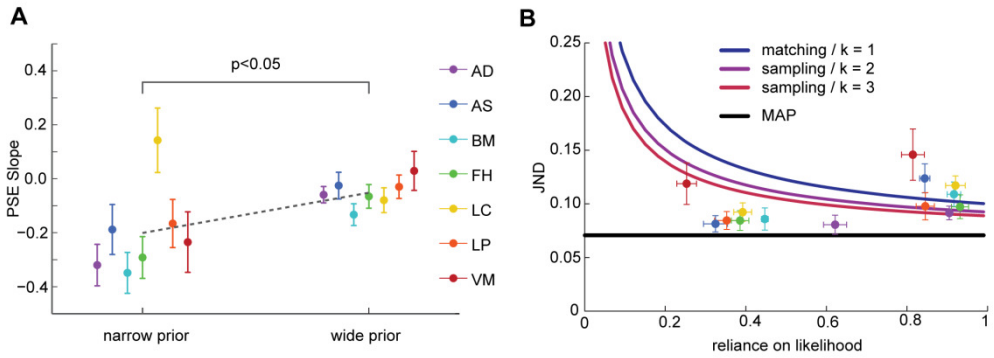


Figure 4.5. Comparing subject data and candidate decision-making theories. A) Across-days average slope of the PSE (during the two-likelihood 2AFC trials) across conditions for each subject. With one exception, each subject's data followed the same trend indicating the prior had a significant effect on their choices during the 2AFC task. B) Each subjects JND plotted versus their reliance on the likelihood (95% CI) Solid lines correspond to the theoretical predictions for the MAP (black line) and sampling (various number of samples, k) hypotheses.

Subjects' JND's did not change with their prior

To summarize, our results so far suggest that subjects learned two distinct priors during the estimation trials, and that they used these priors during the 2AFC trials. We are now in conditions of testing whether the prior uncertainty affects the JND obtained using the standard (same-likelihood) 2AFC task and look for evidence of either decision-making theory.

The MAP hypothesis predicts that changes in the prior should not influence the JND, whereas the sampling hypothesis predicts that changes in the prior do influence the JND – concretely, that a decrease in prior uncertainty should lead to an increase in the JND. We found that subject's JND's did not increase as prior uncertainty (reliance on likelihood) decreases ($p=0.91$, paired t-test, see **Figure 4.5B**), and hence, it does not follow the general trend predicted by sampling. This evidence suggests people do not sample.

Notice that, even though we found no evidence for sampling, in theory subjects could be sampling; recall that in the limit of an infinite number of samples, the predictions for the JND are identical for both strategies. However, given that each splash (likelihood) is displayed for only 25ms and masked immediately after, it is reasonable to expect that the number of samples subjects would be able to sample (if they were using a sampling strategy) would be limited. Related to this, notice that for virtually all subjects the JND (and the 95% CI) in the narrow prior (low reliance on likelihood, leftward points in Figure 5B) is well below the sampling predictions for a few samples.

The paradigm presented here offers a straightforward procedure for using behavioral data to examine theories of decision making under uncertainty. Our findings support the hypothesis that, when making decisions under uncertainty, subjects use a MAP process; after combining a subjective prior and their sensory input they choose the option that is most probable. This finding is important for several reasons. Evidence for a MAP process supports the conventional interpretation of the JND and what it measures, i.e. sensory precision. These findings also argue against a popular proposal for the basis of neural computations based on sampling.

4.4 Discussion

The JND obtained from the 2AFC paradigm is often used to measure sensory uncertainty and hence assumed to be immune to changes in the prior uncertainty. Here we showed how the interpretation of the JND is sensitive to the underlying assumptions about which algorithm the brain uses when deciding between two choices. We did this by explicitly computing the predictions from two prominent decision-making hypotheses; the MAP and the sampling/matching hypothesis. The results of an experiment designed to test these predictions are consistent with the MAP theory and hence argue against the sampling/matching theories.

Previous studies have used the 2AFC paradigm to test assumptions about the underlying computational processes implemented in the brain (Gold and Ding, 2013; Liston and Stone, 2008). In fact, the 2AFC paradigm has been used to study how the prior affects perceptual choices (Liston and Stone, 2008). However, and at the same time, the paradigm is more often used to measure sensory uncertainty without explicitly stating or testing the underlying assumptions about the neural computational processes (Ernst and Banks, 2002; Fetsch et al., 2011; Girshick et al., 2011; Stocker and Simoncelli, 2006; Tassinari et al., 2006); e.g. it has been used to measure the prior uncertainty by factoring out the likelihood uncertainty measured using the JND (Girshick et al., 2011). Many applications for 2AFC assume that the JND measures likelihood uncertainty, our study provides a framework for testing this assumption.

The results from our experiment were not trivially expected. Even though one could argue that MAP is the optimal strategy in a 2AFC task, this does not imply that the brain implements it a MAP algorithm. In fact, several studies support the sampling hypothesis; theoretical work shows that a sampling strategy can be optimal under certain assumptions (Sakai and Fukai, 2008; Vul et al., 2009) and experimental work has also argued for

sampling or matching both in cognitive tasks (Gaissmaier and Schooler, 2008) as well as in perceptual tasks (Battaglia et al., 2011; Wozny et al., 2010), or based on spontaneous neural activity (Berkes et al., 2011).

While our results do not support the matching/sampling hypothesis, we cannot rule out matching/sampling or some other decision-making algorithm in other circumstances. Our paradigm should be extended to tasks that have been suggested to show evidence for sampling and also to tasks often used in studies that use JND as a measure of sensory uncertainty. Another possible caveat is that our task is artificial; we chose it because it is an established paradigm for studying learning and representation of prior uncertainty (Berniker et al., 2010a; Vilares et al., 2012) and because natural/evolutionary/innate priors are harder to change (but see Hosoya et al., 2005). It could be possible, however, that the brain uses a different strategy with natural priors. Future work should adapt this paradigm to a more natural prior/task and see if the results stand.

We used masking to limit the number of samples subjects could take if they were using a sampling strategy. Since in the limit of an infinite number of samples the two theories are undistinguishable (at least using our paradigm), limiting the number of samples would help us distinguishing between the two theories. It has been shown for instance that if samples are costly then decisions based on few samples are optimal (Vul et al., 2009). While we cannot put an exact upper bound on the number of samples subjects can take during the 25ms the image is presented, it is reasonable to expect that masking limits the number of samples humans can take if humans use a sampling decision-making algorithm.

We observed that there is a lower bound on the number of samples subjects could be sampling if they were using a sampling decision-making strategy. We have used this, together with the fact that we are masking to argue against sampling. However, we also observe that the JND measurements

are above the theoretical prediction for MAP. This is explained by the fact that we used the experimentally defined variance of the likelihood. As the true variance of the likelihood is expected to be higher than the experimentally imposed one -- due for instance to sub-optimality in estimating the centroid of the cloud of dots (Tassinari et al., 2006) -- the true JND for a MAP strategy should also be above the theoretical prediction and hence be consistent with our results.

4.5 Materials and Methods

Experimental Protocol

We designed an experiment to examine whether or not subjects' behavior during a 2AFC task is influenced by their prior. If a change in their prior produces systematic changes in their JND, then this would be evidence that decisions are made by sampling. If, on the other hand, the JND is invariant with respect to their prior, then this would offer evidence that subjects use MAP. Additionally, this would provide evidence that the 2AFC task measures sensory precision. To test this, we had subjects participate in a previously published “coin-catching” paradigm (Berniker et al., 2010a; Sato and Aihara, 2011; Vilares et al., 2012) which consists of an estimation task. Here we adapted the paradigm to include 2AFC trials as well as estimation trials.

Estimation and 2AFC trials

A virtual coin-catching paradigm was used to test subjects in both estimation and 2AFC tasks. All trials/tasks began the same. The locations of two virtual coins were drawn from a normal distribution (the prior). The location of the first coin was depicted by quickly presenting a “splash” (the likelihood) as five small red dots drawn from a normal distribution centered on the coin's position (see **Figure 4.2**). After 25 milliseconds a mask (see below) was displayed for 500 milliseconds. Then a second splash was used to depict the location of the second coin (again centered on the coin's location and displayed for 25 milliseconds and followed by a mask. After this subjects were randomly asked to either estimate the second coin's location (the estimation task), or which of the two coins landed further to the right (the 2AFC task) (see **Figure 4.2**).

Estimation task.

In the estimation trials, subjects were presented with a virtual net, depicted with a vertical bar (10% of the screen width). Their task was to place the net where they believed the coin landed. Since the net covered the entire height of the screen, the task was a one-dimensional estimation problem. Once they placed the net in the desired location and depressed the mouse key, the true coin location was displayed to them and the trial ended. If a coin landed within the net it was considered caught. A running tally of the number of coins caught as well as their average distance between the net and the coin was displayed. The estimation trials were used to change the subject's prior belief in coin locations, and its data was used to estimate if that change happened.

2AFC task.

In the 2AFC trials, subjects were instructed to guess which tossed coin, unseen to them, landed further to the right by depressing a key (either 1 for the first coin, or 2 for the second coin). The data collected during the 2AFC trials was used to construct psychometric curves. The data allowed us to both measure the subjects' JND and also to verify that subjects used a prior for the coin's location in both the estimation and 2AFC trials (see below).

By manipulating the variance/uncertainty of the coin's prior – the variance of the distribution from witch the hidden coin's position is sampled -, as well as of the likelihood, we were able to change the subjects' prior, infer if the prior was effectively changed (using the estimation trials) and confirm that subjects used the learned prior in the 2AFC task (using the different-likelihood 2AFC trials). Importantly, since we simultaneously measured subjects' JNDs (using the same-likelihood 2AFC trials, see below) we were able to examine the specific predictions of different decision-making theories; concretely, whether and how subjects' JND changes with changes in variance/uncertainty of their prior.

Experimental details

Seven subjects participated in this study (one female) with an average age of 30.1 ± 7.2 years. Four of the participants were naive to the goals of the experiment, signed consent forms and were paid based on their performance. The remaining three participants are authors of this manuscript (DA, MB and HLF). All experimental protocols were approved by the Northwestern University Institutional Review Board and were in accordance with Northwestern University's policy statement on the use of humans in experiments.

Subjects performed the experiment over five days, participating approximately two hours per day. On each day they were seated in front of a computer monitor (approximately 24 inches (52cm wide, 32.5cm high) in a quiet room. Each subject performed two 1000-trial blocks per day (for a total of 10,000 trials across 5 days). The prior over coin locations switched from block to block, from wide to narrow variance on one day, and narrow to wide variance, on the subsequent day, etc. Both priors were normal distributions with zero mean. The narrow prior had a standard deviation of 4% of screen width while the wide prior had a standard deviation of 20% of screen width. To create the splashes, we used two standard deviations; one was 2.24% of the screen width, the other 8%. Occasionally target coins close to the left and right side of the monitor would have splashes that fell outside the screen limits. In these trials the splash was resampled until all dots were within the screen limits. In half of the trials, the same likelihood was used for both coins (standard deviation of 8% of screen width). In the remaining, randomly drawn trials, one of the coins' splashes used the 2.24% standard deviation, while the other used the 8% standard deviation.

2AFC trials wherein the two coins had the same likelihood standard deviation – denoted *same-likelihood 2AFC* trials – were used to measure subjects' JND. 2AFC trials wherein the two coins had different likelihood standard deviations – denoted *different-likelihood 2AFC* trials – allowed us to verify that subjects used the prior learned during the estimation trials to judge coin locations during the 2AFC trials. Alternatively it could be that subjects were just using the splash's centroid or a different prior (see below); this is an important control because, even though the MAP hypothesis predicts that subjects' behavior is independent of their prior (see **Figure 4.2**), this would also be the predicted behavior under the sampling hypothesis if subjects, while performing the 2AFC task, merely neglected the prior learned during the estimation trials, and instead relied on a different and unchanged prior.

All trials began as described above, and were randomly assigned to be estimation or 2AFC trials. Each block consisted of 500 estimation trials and 500 2AFC trials, in a random order. To assist subjects in learning the coin's prior quickly, the first half of each block was mostly estimation trials (375 estimation trial and 125 2AFC trials) while the second half were mostly 2AFC trials (125 estimation trials and 375 2AFC trials). After the end of the first block, subjects took a brief (3-5 minute) rest before beginning the second block, with a different prior.

At the start of each day, subjects were instructed on how to complete the estimation and 2AFC tasks, from a prepared manuscript. Subjects were told that someone behind them (the exact location not being important) was tossing coins, one at a time, into the pond/screen. In the estimation trials, their task was to try and “catch” the coin by placing a net (the vertical bar) where they believed the unseen coin landed. They were asked to make the average distance between the net and coin as small as possible, while collecting the maximum number of coins. They were also informed that they would be paid based in part on how small this distance was. Though clear to most subjects, it was explained to them that the vertical component of their guess did not matter, as the net spanned the whole height of the screen. For the 2AFC trials they would have to guess which of the unseen coins landed further to the right. Instructions were provided on how to indicate their choice with a key depress. To reduce the influence of uncontrolled cognitive strategies, subjects were also told that that the person throwing coins was not trying to help or hinder their progress, nor reacting to the choices they made.

At the end of each day, their average distance from the hidden coins during the estimation trials was tallied and they were paid a base rate plus an additional bonus for increasingly small errors.

Data Analysis

Our goal was to find a correspondence in the data with the hypotheses we described before. We use simple linear regressions to infer subjects' prior mean and a relative measure of prior variance/uncertainty – denoted *relative reliance of likelihood*, a proxy for subjects' prior uncertainty (see below). We use a cumulative Gaussian psychometric curve to measure JND and PSE. With these inferred pieces of information, we can investigate whether subjects' JND changes when their prior uncertainty changes.

Estimation task

Measuring subject's prior variance/uncertainty. The estimation task is used both to change subjects' prior uncertainty and to measure it. We assume Gaussian distributions and a reasonable cost-function (e.g. minimizing the squared error). Under these assumptions the best way (Bayes' optimal) of combining the two pieces of information (prior and likelihood) is by weighting their means by their relative precision, i.e., their normalized reciprocal variance (see Körding and Wolpert, 2004b; Trommershäuser et al., 2011). This corresponds to the MAP solution:

$$\hat{c} = \frac{\sigma^2}{\sigma^2 + \sigma_c^2}\mu + \frac{\sigma_c^2}{\sigma^2 + \sigma_c^2}c^s$$

where σ and σ_c denote the likelihood and prior variance respectively, and μ and c_s their respective means. Consider the following model:

$$p(\hat{c} \mid c^s) = \phi(\hat{c}; \mathrm{r}_{\mathrm{bias}} + r_{\mathrm{reliance}} c^s, \sigma_{\mathrm{estimation}})$$

where $\phi(x; \mu, \sigma)$ is a normal density function with mean μ and standard deviation σ. By fitting this model to the data and estimating r_{reliance}, i.e., by

computing the slope of the linear regression between the centroid of the splash of dots (proxy for mean of likelihood/sensed cue (Berniker et al., 2010a; Sato and Aihara, 2011; Vilares et al., 2012) and estimated position (using the position of the net) we can estimate a relative measure of each subject's prior uncertainty, r_{reliance}, which we denote by relative reliance on likelihood. If subjects are Bayesian optimal then $r_{\text{reliance}} = \frac{\sigma_c^2}{\sigma^2 + \sigma_c^2}$. Hence by computing the slope of the linear regression between centroid of splash of dots (proxy for sensed cue) and estimated position (position of the net) we can determine a relative measure of subjects' prior uncertainty. Note that if subjects were sampling then, on average, their response would also be optimal but noisier. This means that we can expect $r_{\text{reliance}} = \frac{\sigma_c^2}{\sigma^2 + \sigma_c^2}$, and hence that r_{reliance} to be a relative measure of reliance in likelihood (thus a proxy for prior uncertainty) independently of whether subjects are using a MAP or sampling strategy.

Measuring subject's mean of the prior. We can also infer subjects' mean of the prior using the same data. For that, notice that we can re-arrange the equation above in the following way

$$c^s = \mu + \frac{\sigma^2 + \sigma_c^2}{\sigma_c^2}\left(c^s - \hat{c}\right)$$

We can hence use as estimate of the subjects' mean of the prior, μ , the intercept of linear regression of c^s as a function of $c^s - \hat{c}$.

2AFC task.

The 2AFC trials can be separated into two different kinds; trials where both likelihoods (splashes) had the same standard deviation (regular 2AFC trials), and trials where the standard deviations were different (control 2AFC trials). Trials with equal standard deviations were used to measure subjects' JND and this was done by fitting a regular psychometric function (with lapse correction, see below). Trials with different standard deviations were used to infer whether subjects where using, during the 2AFC trials, the same prior they learned in the estimation trials (see below).

Psychometric function. Psychometric curve fitting for 2AFC tasks finds a relationship between stimuli's discrepancy and the subject's response. Given the uncertainty and noise inherent to the task, the psychometric curve describes the probability of a response given a pair of stimuli. There are multiple functions that can be used to define the probability of response in a psychometric curve. Here we use a cumulative Gaussian function to make it coincide with the theoretical derivations shown in Results and Supp. Information. We could use an alternative functional form for the psychometric curve (e.g. Weibull or logistic) but the cumulative Gaussian function simplifies our exposition substantially.

The psychometric function used in our analysis is

$$p(z=1\,|\,c_1,c_2)=\Phi\left(c_2-c_1\,;\mathrm{PSE},\sigma_{\mathrm{JND}}\right), \qquad 4.4$$

where c_1 is the reference cue, c_2 is the probe cue, and Φ is the cumulative Gaussian function with mean PSE and standard deviation σ_{JND}. We refer to c_2-c_1 as the experiments' cue discrepancy---as opposed to $c_2^s-c_1^s$, the subjects sensed cue discrepancy.

2AFC trials with same likelihood standard deviation. In these trials we are interested in quantifying subjects' JND. We fit the psychometric function in order to determine the $\widehat{\mathrm{PSE}}$ and $\hat{\sigma}_{\mathrm{JND}}$. We use a psychometric curve with lapse, a small variation of the psychometric curve (see below).

Control 2AFC trials; different likelihood standard deviation. In this case, the PSE (see Equations 4.6, 4.8 and 4.10 in Supp. Information, but not the JND see Equations 4.7 and 4.9 in Supp. Information), depends on the absolute position of the cues. The psychometric curve is thus a two-dimensional surface; effectively we get a different psychometric curve for different positions of the reference cue. Importantly, both the MAP and the sampling hypothesis have the same prediction for how changes in prior uncertainty affect how the PSE values depend on the position of the reference cue. Specifically, they predict that the PSE changes linearly with the position of the reference cue c_1 (see Supplemental Information for details):

$$PSE(c_1) = \frac{\sigma_2^2 - \sigma_1^2}{\sigma_1^2 + \sigma_c^2} c_1 \qquad 4.5$$

In the analysis of these 2AFC trials we chose the reference cue, c_1, to be the cue with higher standard deviation ($\sigma_1 > \sigma_2$). Hence the linear term $\left(\sigma_2^2 - \sigma_1^2\right) / \left(\sigma_1^2 + \sigma_c^2\right)$ is expected to be negative and its absolute value is expected to increase as the uncertainty of the prior (σ_c) decreases. Quantifying this linear relationship thus allows us to examine whether or not subjects use the prior learned during the estimation trials to make the 2AFC judgments.

When fitting the psychometric function to data from these trials we assume that PSE equals $\beta_{\mathrm{bias}} + \beta_{\mathrm{ref}} c_1^e$, i.e., we assume a bias term just like in the regular psychometric curve fitting, plus a linear dependence on the

reference cue as predicted in Equation 4.5. Note that if we set $\beta_{\text{ref}} = 0$ we recover the psychometric curve fitting used for the same likelihood standard deviation trials data. As in the simple psychometric function fitting, it is possible to find β_{bias}, β_{ref}, and $\hat{\sigma}_{\text{JND}}$ optimally by framing the problem within a generalized linear model framework.

Psychometric curve with lapse. Equation 4.4 assumes that subjects make no distraction mistakes; given a sufficiently large discrepancy between the cues, the cumulative Gaussian converges on both sides to a 100% discrimination rate. However, to account for occasional mistakes that subjects may produce, we add a lapse parameter that can be interpreted as a small but not negligible change that subjects commit errors and respond randomly, independent of the discrepancy, with λ probability. The psychometric curve can be then modified to accommodate this change as follows

$$p(z=1 \mid c_1^e, c_2^e) = \lambda + (1-2\lambda) \cdot \Phi\left(c_2^e - c_1^e; \text{PSE}, \sigma_{\text{JND}}\right),$$

where now the curve is bounded between $\lambda\%$ and $(1-\lambda)*100\%$ accuracy. We find the parameters jointly by likelihood maximization.

For estimating, for each subject, the slope of PSE, JND, reliance in feedback and mean of the prior, we used only trials after trial 500 for the first day and after trial 100 for the other days. Confidence intervals were computed using 1000 bootstrap samples.

4.6 Supplemental Information

We start by assuming that the prior over cues, $P(c)$, and the likelihoods, $P(s_1 \mid c_1)$, $P(s_2 \mid c_2)$ are Gaussian functions defined in the following way:

Prior: $P(c) = \mathcal{N}(\mu, \sigma_c^2)$

Likelihoods: $P(s_1 \mid c_1) = \mathcal{N}(c_1, \sigma_1^2)$ and $P(s_2 \mid c_2) = \mathcal{N}(c_2, \sigma_2^2)$

Hence, the posterior $P(c_1 \mid s_1)$ is given by:

$$P(c_1 \mid s_1) = P(s_1 \mid c_1)P(c_1) / P(s_1) = \mathcal{N}\left(\frac{\mu\sigma_1^2 + s_1\sigma_c^2}{\sigma_1^2 + \sigma_c^2}, \frac{\sigma_1^2\sigma_c^2}{\sigma_1^2 + \sigma_c^2} \right)$$

analogously for $P(c_2 \mid s_2)$.

MAP Derivations

For maximizing, the subject's response variable, z is a random variable, equal to 1 if the maximizer of $P(c_2 \mid s_2)$ is greater than the maximizer of $P(c_1 \mid s_1)$,

$$z(s_1, s_2) = \begin{cases} 1 & \text{if } \mathrm{E}\left[\mathrm{P}(\mathrm{c}_2|\mathrm{s}_2)\right] \text{-} \mathrm{E}\left[\mathrm{P}(\mathrm{c}_1|\mathrm{s}_1)\right] \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where, for normal distributions,

$$\mathrm{E}\left[P(c_2 \mid s_2)\right] - E\left[P(c_1 \mid s_1)\right] = \frac{\mu\sigma_2^2 + s_2\sigma_c^2}{\sigma_2^2 + \sigma_c^2} - \frac{\mu\sigma_1^2 + s_1\sigma_c^2}{\sigma_1^2 + \sigma_c^2}$$

where for notational simplicity, we refer to this difference as δ. Therefore, a subject's response is determined through the random variable, δ, which is defined by the two random sensory inputs, s_1, and s_2. From $P(\delta \mid s_1, s_2)$, and using $P(s_1 \mid c_1)$ and $P(s_2 \mid c_2)$, we can integrate out s_1 and s_2 to obtain the probability distribution for δ in terms of the two experimental variables, c_1^e, c_2^e.

$$P(\delta \mid c_1^e, c_2^e) = \mathcal{N}\left(c_2^e - c_1^e + \frac{(c_1^e - \mu)\sigma_1^2}{\sigma_1^2 + \sigma_c^2} + \frac{(\mu - c_2^e)\sigma_2^2}{\sigma_2^2 + \sigma_c^2}, \sigma_c^2 \left[\frac{\sigma_1^2}{(\sigma_1^2 + \sigma_c^2)^2} + \frac{\sigma_2^2}{(\sigma_2^2 + \sigma_c^2)^2} \right] \right)$$

We can then compute the probability of a subject's response, given the two cue locations, c_1^e, c_2^e.

$$P\left(z = 1 \mid c_1^e, c_2^e\right) = P\left(\delta \geq 0 \mid c_1^e, c_2^e\right) = \int_0^\infty P\left(\delta \mid c_1^e, c_2^e\right) d\delta$$

$$P(z = 1 \mid c_1^e, c_2^e) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\mu(\sigma_2^2 - \sigma_1^2) + c_2^e(\sigma_1^2 + \sigma_c^2) - c_1^e(\sigma_2^2 + \sigma_c^2)}{\sqrt{2}\sqrt{\sigma_1^2(\sigma_2^2 + \sigma_c^2)^2 + \sigma_2^2(\sigma_1^2 + \sigma_c^2)^2}} \right) \right]$$

this equation, defines the psychometric curve, the probability of the subject responses given the experimentally manipulated cues. First we note a few important features of this curve. In this most general case, where the two sensed cues have different likelihoods, the psychometric curve is a function of the two cues, c_1^e, c_2^e, and cannot be rewritten in terms of the difference between cues $c_2^e - c_1^e$. The psychometric curve is a function of not only of the discrepancy between the two cues, but also their absolute values. To see this, note that the point of subjective equality (PSE) is found when,

$$c_2^e(\sigma_1^2 + \sigma_c^2) + \mu\sigma_2^2 = c_1^e(\sigma_2^2 + \sigma_c^2) + \mu\sigma_1^2 \qquad 4.6$$

and the just-noticeable-difference, or JND is found to be,

$$\sigma_{\text{JND}} = \frac{\sqrt{\sigma_1^2(\sigma_2^2+\sigma_c^2)^2 + \sigma_2^2(\sigma_1^2+\sigma_c^2)^2}}{\sigma_1^2+\sigma_c^2} \qquad 4.7$$

If, however, the two likelihoods have the same variance, i.e., $\sigma_1 = \sigma_2 = \sigma$, then the psychometric curve collapses to the more familiar form:

$$P(z=1 \mid c_1^e, c_2^e) = \frac{1}{2}\left[1+\text{erf}\left(\frac{c_2^e - c_1^e}{2\sigma}\right)\right]$$

where the PSE zero, and the JND is hence $\sigma_{\text{JND}} = \sqrt{2}\sigma$

Sampling derivations

According to the sampling hypothesis, the subject's response is based on two estimates of the mean, $\hat{c}_1^k$ and $\hat{c}_2^k$, computed using k samples from the corresponding posterior distributions, $P(c_1 \mid s_1)$ and $P(c_2 \mid s_2)$, i.e., $\hat{c}_1^k = \frac{1}{k}\sum_{i\in\{1,\ldots,k\}} \hat{c}_1^i$ where $\hat{c}_1^i \sim P(c_1 \mid s_1)$, analogously for c_2.

Hence $\hat{c}_1^k$ and $\hat{c}_2^k$ are random variables and

$$P(\hat{c}_1^k \mid s_1) = \mathcal{N}\left(\frac{\mu\sigma_1^2 + s_1\sigma_c^2}{\sigma_1^2+\sigma_c^2}, \frac{\sigma_1^2\sigma_c^2}{(\sigma_1^2+\sigma_c^2)k}\right)$$

similarly for $P(\hat{c}_2^k \mid s_2)$. A subject's response variable, z, is chosen according to the following rule:

$$z(s_1, s_2) = \begin{cases} 1 & \text{if } \hat{c}_2^k - \hat{c}_1^k \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Just as with the MAP decision, we define a random variable, $\delta = \hat{c}_2^k - \hat{c}_1^k$ which is defined by the two random sensory inputs, s_1, and s_2. Again, we can integrate out s_1 and s_2 using the likelihoods $P(s_1 \mid c_1)$ and $P(s_2 \mid c_2)$ and obtain the probability distribution for δ in terms of the two experimental variables, c_1^e, c_2^e:

$$P(\delta \mid c_1^e, c_2^e) = \mathcal{N}\left(\mu_\delta, \sigma_\delta^2\right)$$

where the mean and variance are defined as:

$$\mu_\delta = c_2^e - c_1^e + \frac{\sigma_2^2(\mu - c_2^e)}{\sigma_2^2 + \sigma_c^2} - \frac{\sigma_1^2(\mu - c_1^e)}{\sigma_1^2 + \sigma_c^2}$$

$$\sigma_\delta^2 = \frac{\sigma_1^2\sigma_c^2}{(\sigma_1^2 + \sigma_c^2)^2} + \frac{\sigma_2^2\sigma_c^2}{(\sigma_2^2 + \sigma_c^2)^2} + \frac{2\sigma_1^2\sigma_2^2\sigma_c^2 + \sigma_c^4(\sigma_1^2 + \sigma_2^2)}{(\sigma_1^2 + \sigma_c^2)(\sigma_2^2 + \sigma_c^2)k}$$

We can then compute the probability of a subject's response, given the two cue locations, c_1^e, c_2^e:

$$P\left(z = 1 \mid c_1^e, c_2^e\right) = P\left(\delta \geq 0 \mid c_1^e, c_2^e\right) = \int_0^\infty P\left(\delta \mid c_1^e, c_2^e\right) d\delta$$

$$P(z = 1 \mid c_1^e, c_2^e) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\mu_\delta}{\sqrt{2}\sigma_\delta}\right)\right]$$

Also in this case, as with the MAP decision hypothesis, the psychometric curve is not translation invariant, i.e. it cannot be written in terms of the

difference between c_1^e , c_2^e . The PSE is, as in the MAP case (Equation 4.6), obtained when

$$c_2^e(\sigma_1^2+\sigma_c^2)+\mu\sigma_2^2 = c_1^e(\sigma_2^2+\sigma_c^2)+\mu\sigma_1^2 \qquad 4.8$$

On the other hand, the just-noticeable-difference is different from the MAP case, given by

$$\sigma_{\text{JND}} = \frac{(\sigma_1^2+\sigma_c^2)}{\sigma_c^2}\sqrt{\frac{\sigma_1^2\sigma_c^4}{(\sigma_1^2+\sigma_c^2)^2}+\frac{\sigma_2^2\sigma_c^4}{(\sigma_2^2+\sigma_c^2)^2}+\frac{2\sigma_1^2\sigma_2^2\sigma_c^2+\sigma_c^4(\sigma_1^2+\sigma_2^2)}{(\sigma_1^2+\sigma_c^2)(\sigma_2^2+\sigma_c^2)k}} \qquad 4.9$$

When the two likelihoods have the same variance, σ then the psychometric curve collapses to,

$$P(z=1\,|\,c_1^e,c_2^e) = \frac{1}{2}\left[1+\text{erf}\left(\frac{(c_2^e-c_1^e)\sigma_c\sqrt{k}}{2\sigma\sqrt{\sigma^2+(k+1)\sigma_c^2}}\right)\right]$$

where the PSE is zero, and the JND is:

$$\sigma_{\text{JND}} = \sqrt{2}\sigma\frac{\sqrt{\sigma^2+(k+1)\sigma_c^2}}{\sigma_c\sqrt{k}}$$

Using the PSE to investigate if subjects transfer the prior to the 2AFC task

When the likelihoods of the two cues displayed during the 2AFC task have different standard deviations – different-likelihood 2AFC trials --, the psychometric curve and, in particular, the PSE, are a function of not only the cue discrepancy but also of the cues' absolute position. For fixed standard deviations of the two cues, we can thus compute the PSE as a function of the reference cue position. We now show that both MAP and Sampling

theories predict the PSE to change linearly with the position of the reference cue, and that this linear relationship depends - in the same way for both theories - on the variance of the prior.

We have seen above (Equations 4.6 and 4.8) that for both the MAP and the Sampling hypothesis the PSE -- by definition the discrepancy $c_2^e - c_1^e$ that makes $P(z=1 \mid c_1^e, c_2^e) = 0.5$ -- is given by:

$$c_2^e(\sigma_1^2 + \sigma_c^2) + \mu\sigma_2^2 = c_1^e(\sigma_2^2 + \sigma_c^2) + \mu\sigma_1^2$$

Let c_1^e be the cue with large variance and chose it to the reference cue. Let c_2^e hence be the probe cue. We can rearrange the previous equation and observe how cue discrepancy $c_2^e - c_1^e$ at the point of subjective equality (PSE) changes linearly with the position of the reference cue c_1^e:

$$\mathrm{PSE}(c_1^e) = c_1^e \frac{\sigma_2^2 - \sigma_1^2}{\sigma_1^2 + \sigma_c^2} \qquad 4.10$$

Importantly we can see how this linear relationship changes with the variance of the prior. This can be used to check whether, during the 2AFC task, subjects are using the prior they have learned in the estimation task; if the linear relationship between PSE and absolute position of the reference cue changes from one learned prior to the other. This way we can exclude an alternative explanation for an eventual absence of change in the JND, specifically, that subjects might be using a different prior in the 2AFC task than the one learned in the estimation task.

5. Saliency and Saccade Encoding in the Frontal Eye Field During Natural Scene

Hugo L. Fernandes, Ian H. Stevenson, Adam N. Phillips, Mark A. Segraves, and Konrad P. Kording

Citation. Fernandes HL, Stevenson IH, Phillips AN, Segraves MA, Kording KP (2013) Saliency and Saccade Encoding in the Frontal Eye Field During Natural Scene Search. Cerebral Cortex doi:10.1093/cercor/bht179

Author Contributions. Conceived and designed the experiments: ANP and MAS. Performed the experiments: ANP and MAS. Analyzed the data: HLF IHS KPK. Implemented the analysis: HLF with the contribution of IHS. Wrote the paper: HLF IHS MAS and KPK

5.1 Summary

The frontal eye-field (FEF) plays a central role in saccade selection and execution. Using artificial stimuli, many studies have shown that the activity of neurons in the FEF is affected by both visually salient stimuli in a neuron's receptive field and upcoming saccades in a certain direction. However, the extent to which visual and motor information is represented in the FEF in the context of the cluttered natural scenes we encounter during everyday life has not been explored. Here we model the activities of neurons in the FEF, recorded while monkeys were searching natural scenes, using both visual and saccade information. We compare the contribution of bottom-up visual saliency (based on low-level features such as brightness, orientation, and color) and saccade direction. We find that, while saliency is correlated with the activities of some neurons, this relationship is ultimately driven by activities related to movement. Although bottom-up visual saliency contributes to the choice of saccade targets, it

does not appear that FEF neurons actively encode the kind of saliency posited by popular saliency map theories. Instead, our results emphasize the FEF's role in the stages of saccade planning directly related to movement generation.

5.2 Introduction

One of the most frequent decisions in our lives is where to look next. How the nervous system makes this decision while free-viewing or searching for a target in natural scenes is an ongoing topic of research in computational neuroscience (Ehinger et al., 2009; Elazary and Itti, 2008; Foulsham et al., 2011; Kayser et al., 2006; Koch and Ullman, 1985; Yarbus, 1967; Zhao and Koch, 2011). The most prominent models of saccade target selection during free-viewing are based on the concept of bottom-up saliency maps. In these models, the image is separated into several channels including color, light intensity, and orientation, to create a set of "feature maps" (for a review see Cave and Wolfe, 1990; Itti and Koch, 2000, 2001b; Koch and Ullman, 1985; Schall and Thompson, 1999; Treisman, 1988). For example, the horizontal feature map would have high values wherever the image has strong horizontal edges. After normalizing and combining these feature maps the output indicates locations in an image containing features that are different from the rest of the image. The more dissimilar an image region is from the rest of the image the more salient or "surprising" it is (Itti and Baldi, 2006). Saliency models for saccade target-selection predict that human subjects are more likely to look at locations that are salient in the sense of being different from the rest of the image. Models based on these ideas have successfully described eye-movement behavior in both humans and monkeys (Berg et al., 2009; Einhäuser et al., 2006; Foulsham et al., 2011). How the brain may implement such algorithms is a central question in eye-movement research.

The involvement of cerebral cortex in this selection of eye movements has been recognized since the late 19th century when David Ferrier reported that eye-movements could be evoked from several regions of the rhesus monkey's cerebral cortex by using electrical stimuli (Ferrier, 1875). One of these regions included an area of cortex now known as the frontal eye field (FEF). The visual and movement-related response field properties of different classes of neurons in the FEF have been carefully characterized using physiological and behavioral methods (Bichot et al., 1996; Bizzi, 1968; Bizzi and Schiller, 1970; Bruce and Goldberg, 1985; Burman and Segraves, 1994; Dias et al., 1995; Dias and Segraves, 1999; Everling and Munoz, 2000; Fecteau and Munoz, 2006; Mohler et al., 1973; Phillips and Segraves, 2010a; Ray et al., 2009; Sato and Schall, 2003; Schall, 1991; Schall and Hanes, 1993; Schall et al., 1995; Schiller et al., 1980; Segraves, 1992; Segraves and Goldberg, 1987; Segraves and Park, 1993; Serences and Yantis, 2006; Sommer and Tehovnik, 1997; Sommer and Wurtz, 2001; Suzuki and Azuma, 1977; Thompson and Bichot, 2005). In addition to eye-movement related activity, FEF firing rates are thought to be affected by simple image features (Peng et al., 2008), by task-relevant features (Bichot and Schall, 1999; Murthy et al., 2001; Thompson et al., 1997; Thompson et al., 1996), and by higher-order cognitive factors including memory and expectation (Thompson et al., 2005). During the period of fixation between saccades, the initial visual activity of FEF neurons is not selective for specific features such as color, shape, or direction of motion (Schall and Hanes, 1993). Later activity, however, is more closely related to saccade target selection, and appears to be influenced by both the intrinsic, bottom-up saliency of potential targets as well by their similarity to the target (Murthy et al., 2001; Thompson and Bichot, 2005; Thompson et al., 2005). Several studies have suggested that the pre-saccadic peak of FEF visual activity specifies the saccade target (Bichot and Schall, 1999; Schall and Hanes, 1993; Schall et al., 1995) and that this visual selection signal is independent of saccade production (Murthy et al., 2001; O'Shea et

al., 2004; Sato et al., 2001; Thompson et al., 1997). In summary, activity in FEF has been linked to information about perception, decision making, planning, and action, increasing the difficulty of identifying a precise computational role for this area.

Although it has been suggested that FEF neurons encode a visual saliency map, the definition for visual saliency in this context is typically largely subjective, non-uniform across studies, not always explicitly defined, and based primarily upon the likelihood that a feature in visual space will become the target for a saccade (Thompson and Bichot, 2005). From a computational perspective, the prevalent interpretation of saliency in the oculomotor field includes the fundamental visual features contributing to the objective definition of visual saliency as well as other factors determining saccade target choice including relevance or similarity to the search target and the gist - the likelihood that the target will be found at a particular location (Itti and Koch, 2000; Land and Hayhoe, 2001; Oliva et al., 2003; Turano et al., 2003). In this study, we use a precise bottom-up definition of saliency, a definition that is independent of task objective or gist, based only upon the basic physical image features. Examining FEF activity in the light of a formal definition will advance our understanding of both the process for saccade target choice as well as the role of the FEF in that process.

It is unclear whether results from earlier studies using artificial stimuli will hold for natural scenes, and exactly how visual and motor information are represented in FEF during naturalistic eye movements. To understand how the brain works ultimately implies understanding how it solves the kinds of tasks encountered during everyday life (Kayser et al., 2004). Following that philosophy, multiple communities have begun to analyze the brain using natural stimuli (Rolls and Tovee, 1995; Sharpee et al., 2004; Smyth et al., 2003; Theunissen et al., 2001; Vinje and Gallant, 2002; Wainwright et al., 2002; Weliky et al., 2003; Willmore et al., 2000) and have quantified the statistics of natural scenes and movements (Bell and Sejnowski, 1997;

Howard et al., 2009; Hyvarinen et al., 2003; Ingram et al., 2008; Lewicki, 2002; Möller et al., 2003; Olshausen and Field, 1996; Schwartz and Simoncelli, 2001; Smith and Lewicki, 2006; Srivastava et al., 2003; Van Hateren and van der Schaaf, 1998). Importantly, these studies show experimentally that surprising nonlinear aspects of processing become apparent as soon as natural stimuli are used (Kayser et al., 2003; MacEvoy et al., 2008; Theunissen et al., 2000). For example, input to regions outside of the classical receptive field during natural scene viewing increases the selectivity and information transfer of V1 neurons (Vinje and Gallant, 2002). It is thus important to analyze FEF activity using natural stimuli.

Although a complete understanding of the FEF's role in eye movement control will depend upon the use of natural stimuli, analyzing neural activities during the search of natural images is a difficult problem. The main factor contributing to this difficulty is the fact that most variables of interest are correlated with one another. It is known that monkeys tend to look at visually salient regions of images (Berg et al., 2009; Einhäuser et al., 2006). This means, that a purely movement related neuron would have correlations with bottom-up saliency. In an extreme example, eye muscle motor neurons responsible for moving the eyes to the right, would on average have more activity during times where the right side of the image has high saliency, and would thus appear to encode high saliency in the right visual field. We clearly would not want to conclude that the motor neuron encodes saliency. This emphasizes the need for a way of dealing with the existing correlations.

To deal with such cases, statistical methods have been developed that enable "explaining away" (Pearl, 1988). In natural scene search, the correlations are not perfect - the subject may not always look to the right when the rightward region is salient. These divergences enable us to identify the relative contributions of visual and motor activation to neuron spiking. The basic intuition is the following: For the case of a neuron that is tuned

only to saliency, its activity may also be correlated with eye movement direction due to the imperfect correlation of saliency and saccade direction during natural scene search. Once we subtract the best prediction based on saliency, however, any correlation with movement would be gone. The opposite would not be true. If we subtract the best prediction based on movement, a correlation with saliency would still exist. Over the past few years, generalized linear models (GLMs), have proven to be powerful tools for solving such problems, modeling spike trains when neural activity may depend on multiple, potentially correlated, variables (Pillow et al., 2008; Saleh et al., 2010; Truccolo et al., 2005).

Here we recorded from neurons in the frontal eye field (FEF) while a monkey searched for a small target embedded in natural scenes. We then analyzed the spiking activity of these neurons using GLMs that treat both bottom-up saliency and saccades as regression variables. Almost all neurons had correlations with upcoming saccades and most also had correlations with bottom-up saliency. However, after taking into account the saccade related activities, the correlations with saliency were explained away. These results suggest that conventional, bottom-up saliency is not actively encoded in the FEF during natural scene search.

5.3 Materials and Methods

The animal surgery, training, and neurophysiological procedures used in these experiments are identical to those reported in (Phillips and Segraves, 2010). All procedures for training, surgery, and experiments were approved by Northwestern University's Animal Care and Use Committee.

Animals and Surgery

Two female adult rhesus monkeys (*Macaca mulatta*) were used for these experiments, identified here as MAS14 and MAS15. Each monkey received

preoperative training followed by an aseptic surgery to implant a subconjunctival wire search coil for recording eye movements (Judge et al., 1980; Robinson, 1963), a Cilux plastic recording cylinder aimed at the frontal eye field (FEF), and a titanium receptacle to allow the head to be held stationary during behavioral and neuronal recordings. Surgical anesthesia was induced with the short-acting barbituate thiopental (5-7 mg/kg IV), and maintained using isoflurane (1.0-2.5%) inhaled through an endotracheal tube. The FEF cylinder was centered at stereotaxic coordinates anterior 25 mm and lateral 20 mm. The location of the arcuate sulcus was then visualized through the exposed dura and the orientation of the cylinder adjusted to allow penetrations that were roughly parallel to the bank of the arcuate sulcus. Both monkeys had an initial cylinder placed over the left FEF. Monkey MAS14 later had a second cylinder place over the right FEF.

Behavioral Paradigms

We used the REX system (Hays et al., 1982) based on a PC computer running QNX (QNX Software Systems, Ottawa, Ontario, Ca), a real-time UNIX operating system, for behavioral control and eye position monitoring. Visual stimuli were generated by a second, independent graphics process (QNX – Photon) running on the same PC and rear-projected onto a tangent screen in front of the monkey by a CRT video projector (Sony VPH-D50, 75Hz non-interlaced vertical scan rate, 1024×768 resolution). The distance between the front of the monkey's eye to the screen was 109.22cm (43 inches).

Visually guided and memory-guided delayed saccade tasks

Monkeys fixated a central red dot for a period of 500-1000 ms. At the end of this period, a target stimulus appeared at a peripheral location. On visually guided trials, the target remained visible for the duration of the trial. On

memory-guided trials, the target disappeared after 350 ms. After the onset of the target, monkeys were required to maintain central fixation for an additional 700-1000 ms until the central red dot disappeared, signaling the monkey to make a single saccade to the target (visually guided) or the location at which the target had appeared (memory-guided). The delay period refers to the period of time between the target onset and the disappearance of the fixation spot. These two tasks were used to characterize the FEF cells by comparing neural activity during four critical epochs (see *Data Analysis*). Typically, trials of these types were interleaved with each other, and with the scene search tasks described below.

Scene search task

This task was designed to generate large numbers of purposeful, self-guided, saccades. Monkeys were trained to find a picture of a small fly embedded in photographs of natural scenes. After monkeys learned the standard visually guided and memory-guided search tasks, the target spot was replaced with the image of the fly. After 30 minutes the scene task was introduced. Both monkeys used in this experiment immediately and successfully sought out the fly. After a few sessions performing this task, it became obvious that monkeys were finding the target after only one or two saccades. We therefore used a standard alpha blending technique to superimpose the target onto the scene. This method allows for varying the proportions of the source (target) and destination (the background scene) for each pixel, and was used to create a semi-transparent target. Even after extensive training, we found that the task was reasonably difficult with a 65% transparent target, requiring the production of multiple saccades while the monkeys searched for the target. Monkeys began each trial by fixating a central red dot for 500-1000 ms, then the scene and embedded target appeared simultaneously with the disappearance of the fixation spot, allowing monkeys to begin searching immediately. The fly was placed

pseudo-randomly such that its appearance in one of eight 45° sectors of the screen was balanced. Within each sector its placement was random between 3 and 30 degrees of visual angle from the center of the screen. Trials ended when the monkeys fixated the target for 300 ms, or failed to find the target within 25 saccades. Images of natural scenes were pseudo-randomly chosen from a library of >500 images, such that individual images were repeated only after all images were displayed. An essential feature of this task is that, although they searched for a predefined target, the monkeys themselves decided where to look. The location where the target was placed on the image did not predict the amplitudes and directions of the saccades that would be made while searching for it nor the vector of the final saccade that captured it.

Image database

The set of images was collected by one of the co-authors (ANP) for the purpose of conducting the experiment in Phillips and Segraves (2010), and is available for download. The photographs were taken using a digital camera, and included scenes with engaging objects such as animals, people, plants, or food. The images were taken by a human photographer and thus may contain biases not present in truly natural visual stimuli (Tseng et al., 2009). For instance, the center of the image tends to be more salient than the edges (as presented in Results, **Figure 5.2A** and **B**).

Neural Recording

Single neuron activity was recorded using tungsten microelectrodes (A-M Systems, Inc., Carlsborg, WA). Electrode penetrations were made through stainless steel guide tubes that just pierced the dura. Guide tubes were positioned using a Crist grid system (Crist et al., 1988, Crist Instrument Co., Hagerstown, MD). Recordings were made using a single electrode advanced by a hydraulic microdrive (Narashige Scientific Instrument Lab,

Tokyo, Japan). On-line spike discrimination and the generation of pulses marking action potentials were accomplished using a multi-channel spike acquisition system (Plexon, Inc., Dallas, TX). This system isolated a maximum of 2 neuron waveforms from a single FEF electrode. Pulses marking the time of isolated spikes were transferred to and stored by the REX system. During the experiment, a real-time display generated by the REX system showed the timing of spike pulses in relationship to selected behavioral events.

The location of the FEF was confirmed by our ability to evoke low-threshold saccades from the recording sites with current intensities of ≤50 μA, and the match of recorded activity to established cell activity types (Bruce and Goldberg, 1985). To stimulate electrically, we generated 70 ms trains of biphasic pulses, negative first, 0.2 ms width per pulse phase delivered at a frequency of 330 Hz.

Data Analysis – General Analysis

FEF cell characterization

We examined average cell activity during four critical epochs while the monkey performed the memory-guided delayed saccade task to determine if the cell displayed visual or pre-motor activity. If not enough data was available from this task, data from the visually guided delayed saccade task was used. The baseline epoch was the 200 ms preceding target onset, the visual epoch was 50-200 ms after target onset, the delay epoch was the 150 ms preceding the disappearance of the fixation spot, and the pre-saccade epoch was the 50 ms preceding the saccade onset. FEF cells were characterized by comparing epochs in the following manner using the Wilcoxon sign-rank test. If average firing rates during the visual or delay epochs was significantly higher than the baseline rate, the cell was considered to have visual or delay activity respectively. If the activity during the pre-saccade epoch was significantly greater than the delay epoch, the

cell was considered to have pre-motor activity. These criteria are similar to those used by Sommer and Wurtz (2000). The selection of neurons for this study was biased towards those with visual activity and our sample does not include any neurons with only motor activity.

IK-Saliency

We considered the Itti-Koch (IK)-saliency (Itti and Koch, 2000; Walther and Koch, 2006) as the definition of saliency (see **Supp. Mat.** and **Figure 5.9A**). This method provides a bottom-up definition of saliency based only on basic image features and independent of task objectives. We used the publicly available toolbox (Walther and Koch, 2006) for computing IK-saliency with the default parameter values and considered three, equally-weighted, channels: color, intensity and orientation. IK-saliency for each image was centered by subtracting the mean of the IK-saliency of that image. To account for a possible imprecision of eye position tracking, we low-pass filtered the IK-saliency using a 5 degree standard deviation 2D-Gaussian (some examples are shown in Results, **Figure 5.2A**). We redid the analysis either without centering or without low-pass filtering the definition of IK-saliency and show that the conclusions of this study are the same (these results are shown in Supplementary Material, **Figure 5.13**).

ROC curve

To compute the Receiver Operator Characteristic (ROC) curve for IK-saliency as an eye fixation predictor we considered all the saccades for both monkeys in the interval between 200ms and 2000ms of each trial. We varied a threshold across the domain of possible values of IK-saliency and determined the fraction of fixations that fell on pixels with IK-saliency above that threshold (y-axis of ROC curve). We compared this true positive rate across all frames to the fraction of pixels without fixations that had IK-saliency above the threshold (the false positive rate). We bootstrapped

across the pixels with fixations to obtain a 95% confidence interval for the area under the ROC curve.

Finally, to test for the predictive value of saliency independent of center-bias we compared, using Mann-Whitney test, the IK-saliency at the fixated locations with the IK-saliency at the same locations in all other images.

Peri/Post-Stimulus time histograms (PSTHs)

We used PSTHs to examine preferred directions for saccades as well as sensitivity to visual saliency. For the saccade-related PSTHs, we considered a time interval of 400 ms centered on saccade onset. We considered all saccades in the time period of 200 ms after trial start and until a maximum of 5000 ms into the trial (less if the trial ended before 5000 ms). We assigned the neuron's activity for each 400 ms perisaccadic interval to one of 8 PSTHs according to the saccade direction and ignoring the magnitude of the saccade. To construct the PSTHs, spikes were binned in 10 ms windows and averaged across trials.

The PSTHs for activity driven by visual saliency were computed in an analogous way. Each of our analyses considers the whole distribution of IK-saliency over the scene to characterize neural responses. We considered a time interval of 400 ms centered on fixation onset. The spikes were binned in 10 ms windows. Activity for a fixation interval was assigned to a particular direction if, after convolving the IK-saliency image with one of 8 filter windows that corresponded to each representative direction relative to eye fixation, the average pixel value was positive. Unlike the saccade-related PSTHs where each raster was associated with only one of the 8 representative directions, IK-saliency for a given image was often elevated in more than one of the 8 filter windows, and thus each raster in the visual saliency PSTHs could be assigned to more than one PSTH. The 8 filter windows were cosine functions of the angle, each with a maximum at the correspondent representative direction and independent of the distance to

fixation point. These filters were thresholded to be zero at a distance smaller than 3 or larger than 60 degrees.

Data Analysis – Generative Model, Model Fitting and Model Comparison

To explicitly model the joint contribution of saliency and saccades we developed a generative model for FEF spiking using a type of generalized linear model (GLM) -- a linear-nonlinear-Poisson cascade model. We specify how these multiple variables can affect neural firing rates and how firing rates translate to observed spikes. We then fit the model to the observed spikes using maximum likelihood estimation (see below).

Generative model

We considered a time interval starting 200 ms after trial start and until a maximum of 5000 ms into the trial. We wanted to examine three hypotheses: spike trains in FEF neurons encode (i) saccade-related (motor) information alone, (ii) bottom-up saliency alone, or (iii) both motor processes and bottom-up saliency. We model spike activity using a Generalized Bilinear Model (Ahrens et al., 2008). We will explain in detail the *joint* model, i.e., the model that considers both saccade and saliency as covariates – candidate predictors of FEF neuron activity. The *saccade only*, *saliency only* and *full-saccade* models are simplifications of this basic model and will be described after. We start by assuming that the conditional intensity (instantaneous firing rate), λ, of a neuron at time t is a function of the eye movements s_m, visual stimuli, s_v, as well as the time relative to saccade onsets, τ_m, and time relative to fixation onsets, τ_v:

$$\lambda(t \mid f, s, \tau, \alpha) = \exp\left(\alpha + f^M(\tau_m) + f^m(\tau_m, s_m) + f^V(\tau_v) + f^v(\tau_v, s_v)\right)$$

We assume that there are two spatio-temporal receptive fields (STRFs) $f^m(.)$ and $f^v(.)$ for motor (saccade) and for visual (saliency) covariates, respectively. To account for possible non-spatially tuned responses (e.g. untuned temporal modulation preceding fixation onset in a saliency encoding neuron or saccade-locked untuned firing rate change) and for the fact that saccades do not have a fixed duration (a histogram of saccade durations is shown in **Figure 5.10B**), we also allow for the possibility of a purely temporal response – independent of direction of saccade or of saliency stimuli – defined by temporal receptive fields (TRFs) at beginning of saccade, $f^M(.)$ and at end of saccade/beginning of fixation, $f^V(.)$. We assume that these STRFs and TRFs combine linearly and, to ensure that the firing rate is positive, the output of this linear combination is then passed through an exponential nonlinearity. To simplify, we assume that the STRFs are space-time separable

$$\lambda(t \mid g,h,s,\tau,\alpha) = \exp\left(\alpha + f^M(\tau_m) + g^m(\tau_m)h^m(s_m) + f^V(\tau_v) + g^v(\tau_v)h^v(s_v)\right)$$

and that both the TRFs and the STRFs are linear in some basis, such that they can be rewritten as a sum of linear and bilinear forms,

$$\lambda(t \mid \mathbf{X},\mathbf{w},\mathbf{b},\alpha) = \exp\left(\alpha + \mathbf{w}_M^\top \mathbf{X}_M(t) + \mathbf{w}_m^\top \mathbf{X}_m(t)\mathbf{b}_m + \mathbf{w}_V^\top \mathbf{X}_V(t) + \mathbf{w}_v^\top \mathbf{X}_v(t)\mathbf{b}_v\right)$$

The vectors $\mathbf{w}_m$ and $\mathbf{w}_v$ define the temporal components of the STRFs, while $\mathbf{b}_m$ and $\mathbf{b}_v$ define the respective spatial components. The parameter α defines the baseline intensity and $\mathbf{w}_m$ and $\mathbf{w}_v$ are the parameters for the purely temporal responses centered at saccade and fixation onset, respectively. These parameters, together with the motor parameters $\mathbf{w}_m$ and $\mathbf{b}_m$, as well as the saliency parameters $\mathbf{w}_v$ and $\mathbf{b}_v$ of the STRF, fully define the neurons firing rate. Notice that the bilinear components of the model are not strictly linear in the parameters unless we consider the temporal components and the spatial components separately.

Finally, we assume that the observed spikes are drawn from a Poisson random variable with this rate:

$$n_{\text{spikes}}(t) \sim \text{Poisson}\left(\lambda\left(t \mid \mathbf{X}, \mathbf{w}, \mathbf{b}, \alpha\right)\right).$$

Hence if $\mathrm{N}_{[t,t+\Delta t[}$ is the number of spikes during the interval $[t, t+\Delta t[$,

$$\mathrm{N}_{[t,t+\Delta t[} = n_{\text{spikes}}(t+\Delta t) - n_{\text{spikes}}(t) \sim \text{Poisson}\left(\int_t^{t+\Delta t} \lambda\left(t \mid \mathbf{X}, \mathbf{w}, \mathbf{b}, \alpha\right) dt\right).$$

We binned the data in $\Delta t = 10$ ms intervals and we assume constant firing rate $\lambda_{[t,t+\Delta t[}$ within each time bin then

$$\mathrm{N}_{[t,t+\Delta t[} \sim \text{Poisson}\left(\lambda_{([t,t+\Delta t[|\mathbf{X},\mathbf{w},\mathbf{b},\alpha)} \Delta t\right)$$

For notational convenience, in the remainder of the Methods t will denote an index representing the time bin $[t, t+\Delta t[$, with $\Delta t = 10$ ms.

Parametrization of the receptive fields

The form of the STRFs depends on how we construct $\mathbf{X}_m$ and $\mathbf{X}_v$, that is, how we parameterize the spatial and temporal components of the saccade and saliency receptive fields. We parameterize the spatial receptive field for saccades by assuming that the activity of the neuron is cosine tuned for saccade direction, i.e., its firing rate is a function of the cosine of the angular difference between the direction of saccade and some fixed direction, the neuron's preferred direction (Georgopoulos et al., 1982; Hatsopoulos et al., 2007). Specifically, for each time index t we define a vector $\mathbf{d}(t)$ as:

$$\mathbf{d}(t) = \left[\cos(\theta(t)) \;\; \sin(\theta(t))\right]^{\top}$$

if a saccade with direction $\theta(t)$ occurred at time index t, otherwise $\mathbf{d}(t) = \mathbf{0}$. For notation convenience we define a matrix $\mathbf{D}(t) = [d_{t,j}]$, where

$d_{t,j} = \mathbf{d}_j(t)$. This matrix incorporates the spatial component of the saccade covariates for each of the spatial basis functions for all time points. Although there is some evidence that FEF neurons may be tuned to saccade magnitude (Bruce and Goldberg, 1985), we focus on directional tuning here, which appears to be the dominant factor. The construction of $\mathbf{X}_M$ and $\mathbf{X}_V$, which defines the form of the TRFs, is done in an analogous way by defining vector $\mathbf{d}(t)$ as 1 (one dimensional) if a saccade occurred at time index t and 0 otherwise. To model temporal variation near the time of saccade and fixation onset we parameterized temporal receptive fields with a set of 5 basis functions. Since saccades and fixations are defined by very specific points in time, we restrict ourselves to finite windows 200 ms before to 300 ms after saccade or fixation onset. Specifically, our set of temporal basis functions are 5 truncated Gaussians with standard deviation of 50 ms:

$$\mathbf{f}_k(t) = \mathcal{N}\left(\mu_k, 50\text{ ms}\right) \times \mathbf{1}_{[-200,300]}(t),$$

where $\mathbf{1}_A(.)$ is the indicator function, and the k =5 means μ_k are equally spaced such that they partition the interval between -200 ms and 300 ms into 6 subintervals: $\mu_k = [-200 + \Delta\tau : \Delta\tau : 300 - \Delta\tau]$ ms with $\Delta\tau$ =500/6 ms. We incorporate temporal information by convolving each column of matrix $\mathbf{D}$ with each basis function that parameterizes the temporal receptive fields. Finally, we define, for each time index t, the matrix $\mathbf{X}_m(t) = [x_{k,j}(t)]_{5\times 2}$ where

$$x_{k,j}(t) = \left(\mathbf{D}_j * \mathbf{f}_k\right)(t).$$

The matrix $\mathbf{X}_M(t)$ is defined in an analogous way and is hence a 5-by-1 matrix.

For the visual/saliency basis we assume a similar model where the neuron has a preferred direction for the saliency surrounding the eye fixation position. This model is based on the entire saliency distribution across the scene. The parameterization is analogous to the saccade spatial receptive field:

$$\mathbf{d}(t) = \left[\sum_{x,y} \frac{x}{\left| x^2 + y^2 \right|} \mathrm{IKS}\left(x - x_{\mathrm{fix}}(t), y - y_{\mathrm{fix}}(t) \right) \; \sum_{x,y} \frac{y}{\left| x^2 + y^2 \right|} \mathrm{IKS}\left(x - x_{\mathrm{fix}}(t), y - y_{\mathrm{fix}}(t) \right) \right]^{\top}$$

if a fixation started at time t and $\mathbf{d}(t) = 0$ otherwise. IKS denotes the IK-saliency of the current image and the sum is over all pixel positions (x, y) in a window centered at the eye position during fixation (see Supp. Mat. and the description above and (Itti and Koch, 2000; Walther and Koch, 2006) for details). Similarly to the cosine tuning to saccade direction used above, this representation provides a directional tuning to average saliency. We considered the median eye position during the fixation period as the value of eye position during the i$^{\mathrm{th}}$ fixation, $\left(x_{\mathrm{fix}}(t), y_{\mathrm{fix}}(t) \right)$. The construction of matrices $\mathbf{X}_v$ and $\mathbf{X}_V$ is then analogous to the construction of $\mathbf{X}_m$ and $\mathbf{X}_M(t)$.

The STRFs for saccades and saliency allow us to model directional dependence that is then modulated by an envelope around the time of saccade or fixation onset. Note that the STRFs for saliency and saccades are allowed to be completely unrelated under this model, and the same is true for the TRFs, the purely temporal responses around saccade and fixation onset. The joint model has a total of 25 parameters: α for the baseline (1), $\mathbf{w}_M$ and $\mathbf{w}_V$ for the TRFs (5+5), $\mathbf{w}_m$ and $\mathbf{w}_v$ for the temporal response of the STRFs (5+5) and $\mathbf{b}_m$ and $\mathbf{b}_v$ for spatial component of the STRFs (2+2). In addition to the joint model we consider a *saccade-only*

model ($\mathbf{w}_V = 0$, $\mathbf{w}_v = 0$ and $\mathbf{b}_v = 0$, 13 parameters) and a *saliency-only* model ($\mathbf{w}_M = 0$, $\mathbf{w}_m = 0$ and $\mathbf{b}_m = 0$, 13 parameters). Finally we consider also the *full-saccade* model ($\mathbf{w}_v = 0$ and $\mathbf{b}_v = 0$, 18 parameters) which can account for saccade duration variability and some possible temporal representation of the end of the saccade.

Fitting algorithm

To estimate the parameters α , $\mathbf{w}_M$, $\mathbf{w}_m$, $\mathbf{w}_m$, $\mathbf{w}_V$, $\mathbf{w}_v$, $\mathbf{b}_m$ and $\mathbf{b}_v$, we use maximum likelihood estimation and coordinate ascent. By coordinate ascent we mean that we alternate between fitting one subset of parameters and another. We do this because the model is linear only when we consider the temporal and spatial parameters of the bilinear terms separately. We first fit the baseline, the purely temporal parameters and the temporal parameters of the bilinear terms holding spatial parameters fixed, which reduces the problem to a GLM:

$$\lambda(t) = \exp\left(\alpha + \mathbf{w}_M^\top \mathbf{X}_M(t) + \mathbf{w}_m^\top \mathbf{X}_{m,\hat{\mathbf{b}}_m}(t) + \mathbf{w}_V^\top \mathbf{X}_V(t) + \mathbf{w}_v^\top \mathbf{X}_{v,\hat{\mathbf{b}}_v}(t)\right)$$

where $\hat{\mathbf{b}}_m$ and $\hat{\mathbf{b}}_v$ are fixed parameters for the spatial receptive field and $\mathbf{X}_{m,\hat{\mathbf{b}}_m}(t) = \mathbf{X}_m(t)\hat{\mathbf{b}}_m$ and $\mathbf{X}_{v,\hat{\mathbf{b}}_v}(t) = \mathbf{X}_v(t)\hat{\mathbf{b}}_v$. We then repeat the procedure and fit the baseline, the purely temporal parameters and the spatial parameters holding the temporal parameters of the bilinear terms fixed:

$$\lambda(t) = \exp\left(\alpha + \mathbf{w}_M^\top \mathbf{X}_M(t) + \mathbf{X}_{m,\hat{\mathbf{w}}_m}(t)\mathbf{b}_m + \mathbf{w}_V^\top \mathbf{X}_V(t) + \mathbf{X}_{v,\hat{\mathbf{w}}_v}(t)\mathbf{b}_v\right)$$

where $\hat{\mathbf{w}}_m$ and $\hat{\mathbf{w}}_v$ are fixed parameters for the temporal response of the spatially modulated component of the receptive field and $\mathbf{X}_{m,\hat{\mathbf{w}}_m}(t) = \hat{\mathbf{w}}_m^\top \mathbf{X}_m(t)$ and $\mathbf{X}_{v,\hat{\mathbf{w}}_v}(t) = \hat{\mathbf{w}}_v^\top \mathbf{X}_v(t)$.

We alternate between fitting one set of parameters and the other until the log-likelihood converges. Since both likelihood functions are log-concave it is reasonable to expect that it converges to the optimal solution (Ahrens et al., 2008), and, in practice, random restarts converge to the same STRF solutions.

Model comparison

To compare the joint model, the saccade-only, and saliency-only models we computed, using 10-fold cross-validation, the *pseudo R^2* for each model (Haslinger et al., 2012; Heinzl and Mittlböck, 2003) and the *relative pseudo R^2*. Note that we should not use the traditional R^2 to quantify the spike prediction accuracy of the model since while that measure assumes Gaussian noise, the number of spikes is non-negative and discrete signal. Instead we use an extension of the traditional R^2 measure to Poisson distributions; the pseudo R^2. The pseudo R^2 can be interpreted as the relative reduction in deviance due to the additional covariates a model and is defined as:

$$R_D^2(\text{model}) = 1 - \frac{\log \mathrm{L}(n) - \log \mathrm{L}(\hat{\lambda})}{\log \mathrm{L}(n) - \log \mathrm{L}(\bar{n})}$$

where $\log \mathrm{L}(\hat{\lambda})$ is the log-likelihood of the model under consideration, $\log \mathrm{L}(n)$ is the likelihood of the saturated model and $\log \mathrm{L}(\bar{n})$ is the likelihood of the homogenous model. The homogeneous model is the model that assumes a constant firing rate, specifically, the average firing rate of the training set. The saturated model provides an upper-bound on prediction accuracy by assuming that the firing rate in a certain time bin is exactly equal to the observed firing rate in that time bin.

In order to compare between models 1 and 2, where model 1 is a model nested in model 2 – for example, the saccade-only model is nested in the joint model - we use the relative pseudo R^2 which is defined analogously:

$$R_D^2(\text{model 2, model 1}) = 1 - \frac{\log \mathrm{L}(n) - \log \mathrm{L}(\hat{\lambda}_2)}{\log \mathrm{L}(n) - \log \mathrm{L}(\hat{\lambda}_1)}$$

Where $\log \mathrm{L}(\hat{\lambda}_1)$ and $\log \mathrm{L}(\hat{\lambda}_2)$ are the log-likelihood of models 1 and 2, respectively. The relative pseudo R^2 can hence be interpreted as the relative reduction in deviance due to the extra set of covariates included in model 2. Note that $R_D^2(\text{model 1}) = R_D^2(\text{model 1, homogeneous model})$.

It is important to recognize that we are not able to obtain unbiased variance estimates for the pseudo-R^2 obtained using 10-fold cross-validation since the correlations due to the overlap of the testing sets typically leads to underestimating the variance (Bengio and Grandvalet, 2004). However, by bootstrapping across the whole population of recorded neurons and within each subpopulation of visuomotor and visual neurons, we can obtain 95% confidence intervals on the average pseudo R^2 for each population and sub-population of neurons.

There are other measures that we could have used such as *bits per spike* (Harris et al., 2003; Pillow et al., 2008) which is defined as the log (base 2) of the likelihood ratio between the model and the baseline model, divided by the number of spikes. The bits per spike measure gives the reduction in entropy (mutual information) due to the covariates. The pseudo R^2 measure that we use is, apart from the different basis of the logarithm, the bits per spike measure normalized by the amount of bits per spike of the saturated model. Hence the two measures are closely related. It is important to note that, although the pseudo-R^2 measure has the advantage of being upper-bounded by 1, this bound is impossible to achieve in practice unless every spike is perfectly predicted.

Overfitting

We checked for overfitting for every neuron considering all trials and for a particular neuron (neuron 4) as a function of the number of trials. We

computed, for the joint model and for a particular neuron (neuron 4), the pseudo R^2 on test data and on training data as a function of the number of trials used in the analysis. For each set of trials considered we randomly partitioned the data into 10 subsets. We fitted STRFs for all combinations of 9 subsets (the training set) of this partition and computed the pseudo R^2 on the training set and on the remaining 10% (the test set). Finally we computed the mean across 10-folds and obtained an average of spike prediction accuracy on test data and on training data. To check for overfitting for all neurons we repeated the same procedure for every neuron considering all trials.

Simulations

To verify the ability of the model to dissociate saliency and saccade-related spiking we simulated 3 typical kinds of neurons, saccade only, saliency only and joint dependence. We used the behavioral data from one particular neuron in our dataset to simulate spikes, assuming the same STRF for this set of simulations (as presented in Results, **Figure 5.7A**), and using smoothed IK-saliency as the definition of saliency. We assumed that each model had the same saccade/saliency STRF. This STRF was obtained by fitting the saccade-only model to the data of a particular neuron (neuron 4). To compute confidence intervals for recovered angle and recovered temporal filter we split the dataset into a partition of 10 sets of trials with an equal amount of trials. We computed the pseudo R^2 confidence interval using 10-fold cross-validation.

For the next set of simulations (as presented in Results, **Figure 5.8**) we used data from a particular neuron (neuron 4) and we fitted the receptive fields using the saccade only model. We then used baseline and the STRF terms to simulate spike data for a new set of simulated neurons. We tested how adding Gaussian white noise to the IK-saliency affected how well we could recover saliency encoding (as measured by relative pseudo R^2

between the joint model and the saccade only model). We matched the variance of the noise to the variance of the IK-saliency image. Finally, we simulated neurons that lie in the range between only saccade encoding neurons and neurons that encode equal amounts of saccade and saliency, and again tested how well we could recover saliency encoding.

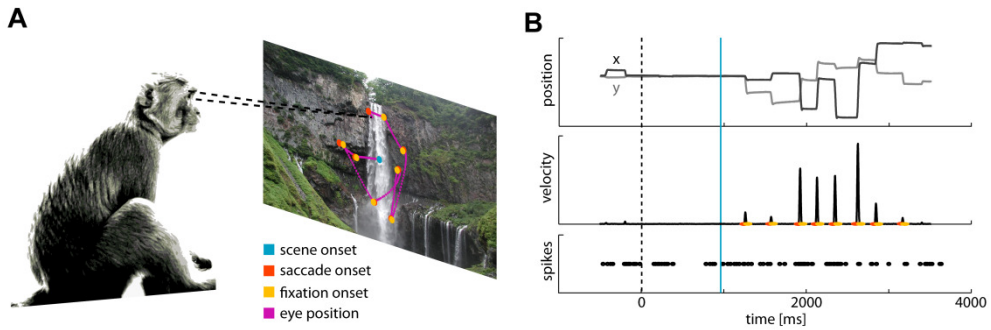


Figure 5.1. Behavioral task and data from a typical trial. (*A*) Monkeys were rewarded for finding the picture of a fly (not shown) embedded in natural scenes. (*B*) Eye position and spike trains were recorded for each trial, allowing us to model dependencies between image features, eye movement, and neural responses. Vertical dashed line marks beginning of fixation of a dot appearing at the center of the tangent screen. Blue vertical line marks the appearance of the image with embedded target. Red and yellow dots mark the beginning and end of saccades. Saccade endpoints correspond to the beginning of a new period of fixation between saccades.

5.4 Results

We recorded from single neurons in the frontal eye field (FEF) of behaving monkeys while they searched for a small inconspicuous target embedded in a natural image stimulus (**Figure 5.1**, target not shown, see Methods). Eye movements where monitored and the monkey was rewarded with water for successfully finding the target. In the following analysis we examine the activity of 52 FEF neurons recorded from 2 rhesus monkeys (MAS14, n =

30; MAS15, n = 22) categorized, using visual and memory-guided saccade tasks, as visual (n = 37) or visuomovement (n = 15) neurons; visual neurons have strong responses after target onset in the receptive field and visuomovement neurons are visual neurons that also have strong activity during the pre-saccade epoch (see Methods - FEF cell characterization). A previous study examined saccade tuning in these data, ignoring visual information (Phillips and Segraves, 2010). Here we analyze how the activity of the neurons relates to aspects of both saccades and features of the natural scene stimuli, more specifically to a bottom-up definition of saliency.

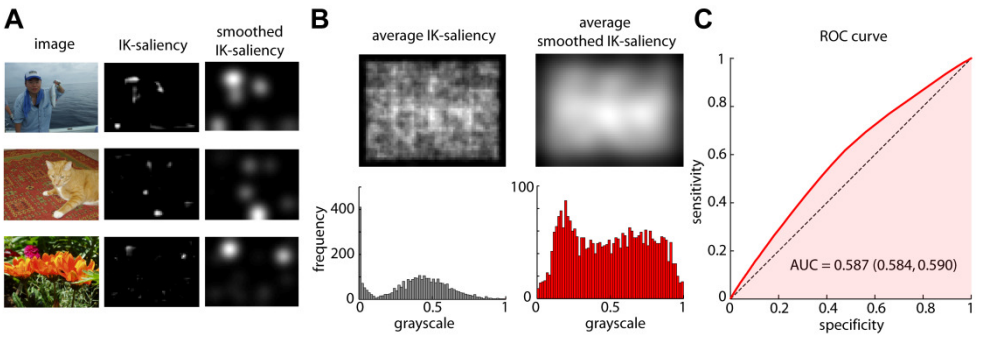


Figure 5.2. Saliency maps and saccade prediction. (*A*) Three typical images from the natural scene search task, along with their IK-saliency maps, and smoothed IK-saliency maps (filtered with an isotropic Gaussian with a standard deviation of 5 deg). (*B*) (top) Average IK-saliency map across all images used in the task along with the average smoothed IK-saliency. Note that there is a bias towards the center of the image being more salient than the edges. (bottom) Corresponding image histograms. (*C*) ROC curve for smoothed IK-saliency as an eye fixation predictor. Area under the curve (median and 95% confidence interval, boostrap) is shown.

We use the definition of saliency (IK-saliency) developed by Itti and Koch (Itti and Koch, 2000). The IK-saliency is a traditional, bottom-up saliency map algorithm that converts images into saliency maps based upon color, intensity and orientation on multiple spatial scales (see (Itti and Koch,

2000), Methods and Supp. Mat. for details). For each of the maps, the algorithm computes how different each location or pixel is from its surround, and the map is then normalized. This leads to conspicuity maps which are then added together to define the overall saliency map. Points that are similar to the rest of the image will have low saliency while, potentially interesting points that are different from the rest of the image have high saliency (**Figure 5.2A** and **B**). The resulting saliency map tends to be highly sparse with most regions of the image being unsurprising (**Figure 5.2A**). There is a non-negligible center bias where the center of the image is more salient than the borders (**Figure 5.2B**), an effect that is due to human photographers having a bias in their choice of pointing direction (Tseng et al., 2009). Saliency maps summarize the high dimensional properties of an image with a single dimension; the saliency or interestingness of the image as a function of space.

We first wanted to check if, as predicted by previous publications (Berg et al., 2009; Einhäuser et al., 2006), monkeys look more often at regions of the image that have high saliency. We thus plotted the standard ROC curve which quantifies how well the saccade targets can be predicted from the saliency map (**Figure 5.2C**). We found the area under the ROC curve to be 0.587 (0.584, 0.590) (median and 95% confidence interval, bootstrap, see Methods for details) – somewhat lower than in previous monkey free-viewing saccade experiments but above the chance level of 0.50 (Einhäuser et al., 2006). However, in our experiment the monkey was not free-viewing the images but had a specific task: it was searching for an embedded target. This top-down goal likely makes the saccades less predictable compared to the case when only bottom-up saliency information is considered. To test if the predictive values of saliency were only due to the center bias, we compared the saliency of image locations where fixations occurred with the average saliency for that location for the remainder of the image set. We found that saliency at the fixated locations

tends to be higher than at the same location in the other images ($p<10^{-10}$, Mann-Whitney test), demonstrating that the predictive value of saliency for fixation choice is due to more than just center-bias. Algorithms that calculate bottom-up saliency predict some aspects of fixation behavior but tend to be somewhat imprecise. When the task is not a free-viewing task but involves target search the predictions of bottom-up saliency maps become even more imprecise. Regarding attempts to understand how saliency relates to the activities of FEF neurons, many methods such as post/peri-stimulus time histograms (PSTH) rely on a well-controlled stimulus or trigger. For those methods, it would be advantageous if saliency did not predict eye movements. Here we use a model-based, multivariate regression approach where saliency and eye movements are not required to be independent.

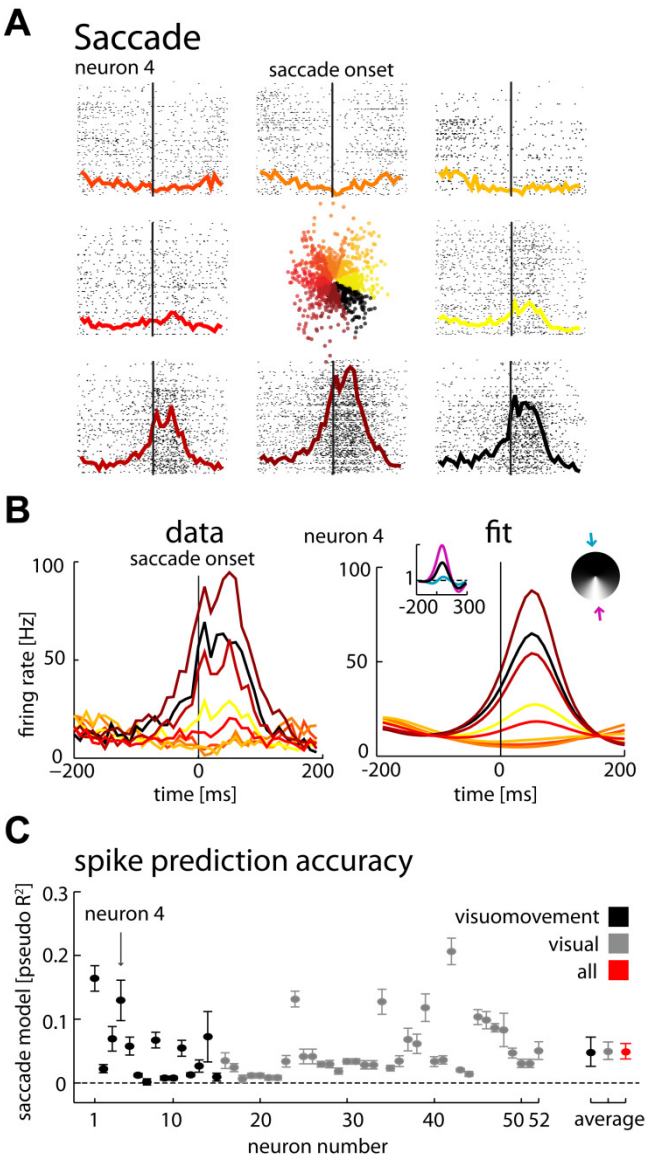


Figure 5.3. Saccade encoding. (*A*) Rasters sorted by direction of saccade, centered on saccade onset and the correspondent peri-stimulus time histograms (PSTHs) for a particular neuron (neuron 4). (B) Overlapping colored PSTHs (left), the fitted spatial and temporal receptive fields (right, insets) and correspondent reproduced PSTHs (right). Actual PSTHs were constructed using all the trials. Parameters were fitted to randomly chosen 90% of the trials and fitted PSTHs were constructed using those 90% of the trials. Blue and purple curves (right, inset) correspond to the temporal gains in the directions of lower (blue arrow) and higher (purple arrow)

modulation of the spatio-temporal receptive fields (STRFs). (*C*) Spike prediction quality for each neuron: Pseudo R^2 (±2SEM, 10-fold cross-validation for each individual neuron; 95% bootstrap confidence intervals for the averages across the recorded population and subpopulations) of the saccade encoding model. Neurons previously classified as visuomovement and visual and respective averages (95% CI, bootstrap across neurons) are shown in black and grey, respectively (see Methods). Global average (95% CI, bootstrap across neurons) is represented in red. Arrow signals neuron number 4, the example neuron in panels A and B.

Saccade Representation

One of the well-established characteristics of many FEF neurons is that they are tuned to the direction of upcoming movements. To quantify this dependence on saccade direction, and test if it may be affected by search in natural images, we estimated each neuron's spatio-temporal tuning to direction of movement. The saccade-triggered PSTH for eye-movements to various octants shows that, indeed, some neurons do have substantial tuning to the direction of saccade (**Figure 5.3A**).

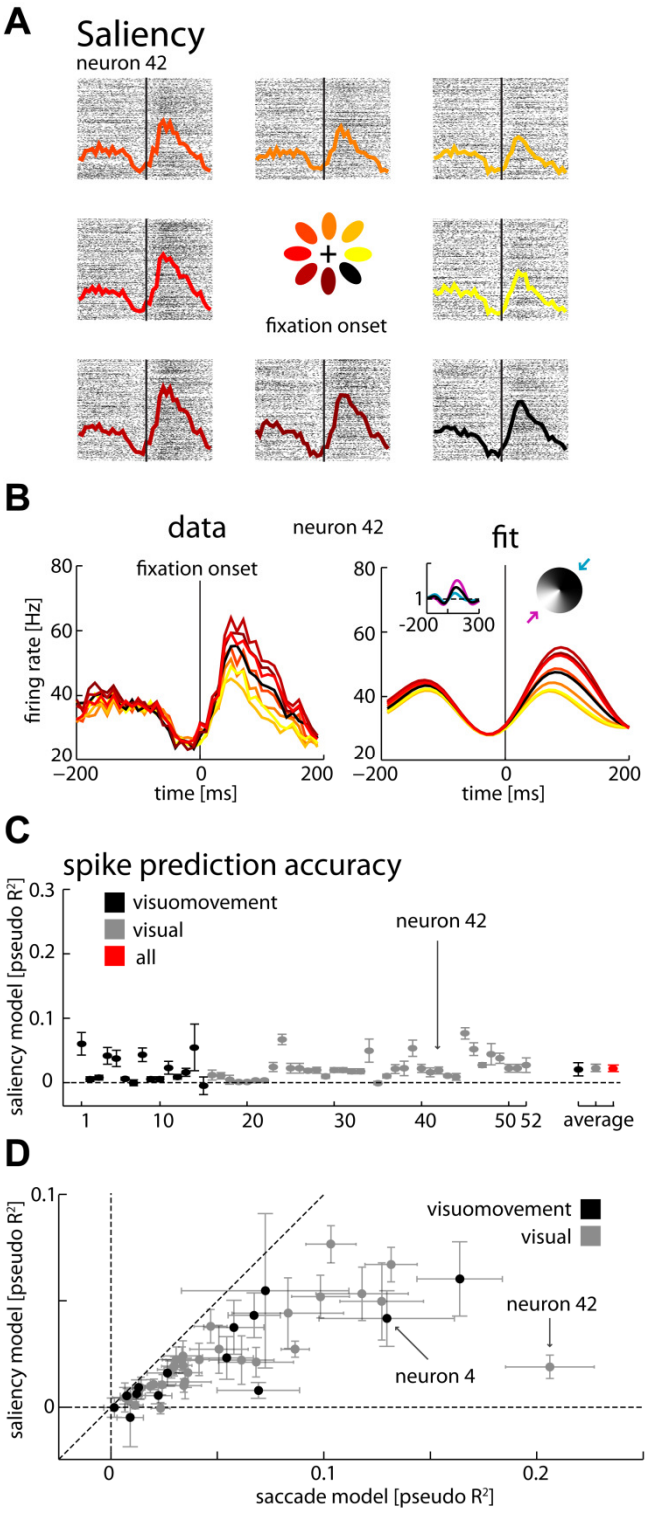

A
Saliency
neuron 42
fixation onset
B
data
neuron 42
fit
fixation onset
firing rate [Hz]
time [ms]
C
spike prediction accuracy
visuomovement
visual
all
neuron 42
saliency model [pseudo R²]
average
D
visuomovement
visual
neuron 42
neuron 4
saliency model [pseudo R²]
saccade model [pseudo R²]

Figure 5.4. Saliency encoding. (*A*) Rasters and post-stimulus time histograms (PSTHs) for a particular neuron (neuron 42). Data are aligned on fixation onset, and assigned to a raster/PSTH based upon the directions where IK-saliency was elevated during the fixation interval (see Methods for additional detail and **Figure 5.11** for the analogous saccade-onset PSTHs for this neuron). (*B*) Overlapping colored PSTHs (left), the fitted spatial and temporal receptive fields (right, insets) and correspondent reproduced PSTHs (right). Actual PSTHs were constructed using all the trials. Parameters were fitted to randomly chosen 90% of the trials and fitted PSTHs were constructed using those 90% of the trials. Blue and purple curves (right, inset) correspond to the temporal gains in the directions of lower (blue arrow) and higher (purple arrow) modulation of the STRFs. (*C*) Spike prediction quality for each neuron: Pseudo R^2 (± 2SEM, 10-fold cross-validation for each individual neuron; 95% bootstrap confidence intervals for the averages across the recorded population and subpopulations) for the vision/saliency encoding model. Neurons previously classified as visuomovement and visual and respective averages (±2SEM) are shown in black and grey, respectively (see Methods). Global average (±2SEM) is represented in red. Arrow signals neuron 42, the example neuron in panels A and B. (*D*) Scatter plot for spike prediction quality (±2SEM, 10-fold cross-validation) of saccade model (same data as **Figure 5.3C**) and saliency (same data as **Figure 5.4C**) for each neuron.

We then used a generalized linear model (GLM, see **Figure 5.9B** and Methods) to explicitly model the spatio-temporal tuning to saccade direction of the neurons. The model used here (space-time separable STRF with cosine direction dependence) accurately captures the properties of the example neuron (**Figure 5.3B**), and allows us to quantify how well-tuned each neuron is to saccades in each direction (**Figure 5.3C**). Most of the neurons we recorded from appear to have strong saccade-related modulation, similar to previous descriptions of neurons in the FEF during simple visual tasks (e.g. Bruce and Goldberg, 1985).

Vision/Saliency Representation

We next wanted to see if the same neurons might also be tuned for visual saliency. Using fixation-triggered PSTHs divided by the direction with the highest IK-saliency, we found that, indeed, some neurons seem to have substantial tuning to directions in which there are salient stimuli (**Figure 5.4A**, but see below). Similar to the saccade direction dependence shown above, we found that a GLM based on tuning to IK-saliency accurately captured the properties of this neuron (**Figure 5.4B**) and allowed us to quantify how well each neuron was tuned to the saliency of the stimuli (**Figure 5.4C**). Using this saliency model, it appears that some of the neurons we recorded from do have significant tuning for salient stimuli in a particular direction.

Explaining Away Saliency Representation

So far we have found that some neurons do appear to have tuning to saccade and also tuning to the direction in which there are salient stimuli. For most neurons, saccade direction alone provides a better model of spiking than saliency alone (**Figure 5.4D**), however, since these two variables are correlated, the independent analyses above may be confounded. We have shown that monkeys tend to make saccades towards more salient targets, even during natural scene search. This means, that if FEF neurons encode only saccade movement, their activity might still be correlated with saliency. Furthermore fixation onset times and saccade onset times are also highly correlated, which may make it difficult to disambiguate the effects of saccades and saliency on spiking activity.

We thus implemented a GLM that predicts spikes based on saccade and saliency at the same time. This approach allows us to take advantage of a statistical effect called explaining away. If the spikes could be fully described by saliency then the system would put no weight on saccade and

vice versa. Since the saccades and saliency are not *perfectly* correlated, such a joint model will determine which of the two factors is, statistically, a more direct explanation of a neuron's firing.

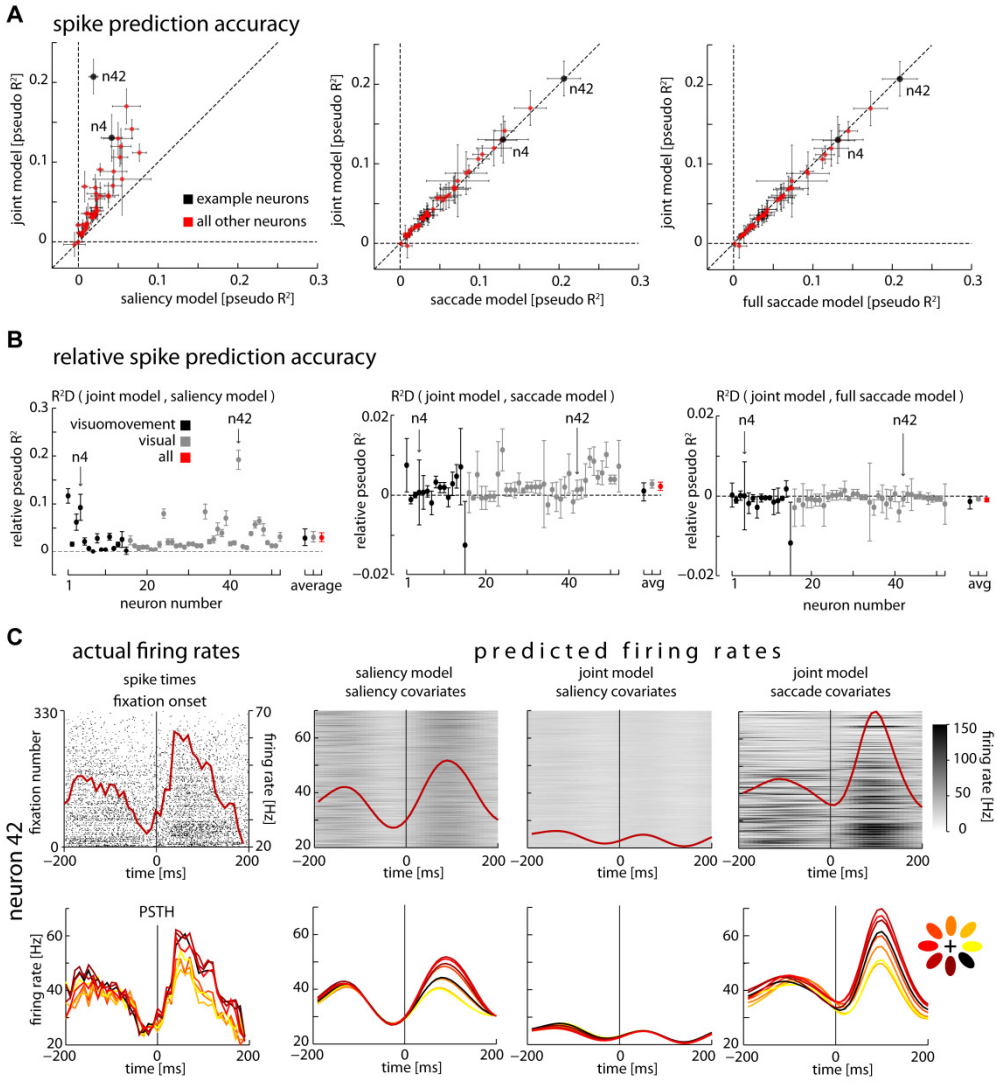


Figure 5.5. Explaining away. (*A*) Scatter plots of the spike prediction accuracy (±2SEM, 10-fold cross-validation, see Methods) under the saliency-only (left)/ saccade-only (center)/ full-saccade (right) and joint models. The saccade-only, saliency-only and full-saccade models are represented on the x-axis and the joint

saccade in the y-axis. “n4” and “n42” denote neurons 4 and 42, the example neurons of **Figure 5.3** and **5.4**, respectively. (*B*) Relative pseudo R^2 between the joint model and the saliency model (left)/ saccade model (center)/ full saccade model (right) (±2SEM, 10-fold cross-validation for each individual neuron; 95% bootstrap confidence intervals for the averages across the recorded population and subpopulations). Arrows signal neurons 4 and 42, the example neurons of **Figure 5.3** and **5.4** respectively. Note different y-axis scales for left versus center and right panels. (*C*) Actual spikes and PSTHs (1st column) and predicted firing rates and PSTHs for saliency only model (2nd column), joint model using saliency covariates only (3rd column) and joint model using saccade covariates only (4th column) for the example neuron of **Figure 5.4** (neuron 42). Parameters were fitted to 50% of the trials and the data shown (both actual spikes and predicted firing rates) correspond to spikes and covariates of the remaining 50% of data (testing set). Upper panels show raw data and predicted firing rates from 340 fixations of the test set where the IK-saliency in the lower-left octant area of the image relative to the point of fixation was positive (see Methods). Lower panels show PSTHs for all directions.

For essentially all of the recorded neurons, we find that adding a spatio-temporal saliency receptive field to the saccade model does not improve the spike prediction accuracy (**Figure 5.5**). A model that uses only saccade and a model that uses both saccade and saliency perform almost equally well (**Figure 5.5A** and **B**, center panels) - in contrast, considering both saccade and saliency improves the performance relative to considering only saliency (**Figure 5.5A** and **B**, left panels). In fact, the apparent saliency related modulation (**Figure 5.4A**, **B** and **Figure 5.5C**, 1st and 2nd columns) can be reproduced using motor information only (**Figure 5.5C**, 4th column). Saccade covariates of the joint model can capture the trial-by-trial variability better than the saliency only model which just smears the spiking activity (**Figure 5.5C**, 4th and 3nd columns, respectively). As saccade duration has some variability (see **Figure 5.10B**) we tested a GLM that adds a purely temporal response centered at the end of the saccade to the saccade-only model: the full-saccade model. We find that it completely explains away –

for all neurons – the saliency modulation (**Figure 5.5A** and **B**, right panels) – the saliency related covariates do not add any predictive power to the full-saccade model. In other words, when modeling activities carefully, there is absolutely no sign of bottom-up saliency (Itti and Koch, 2000) encoding.

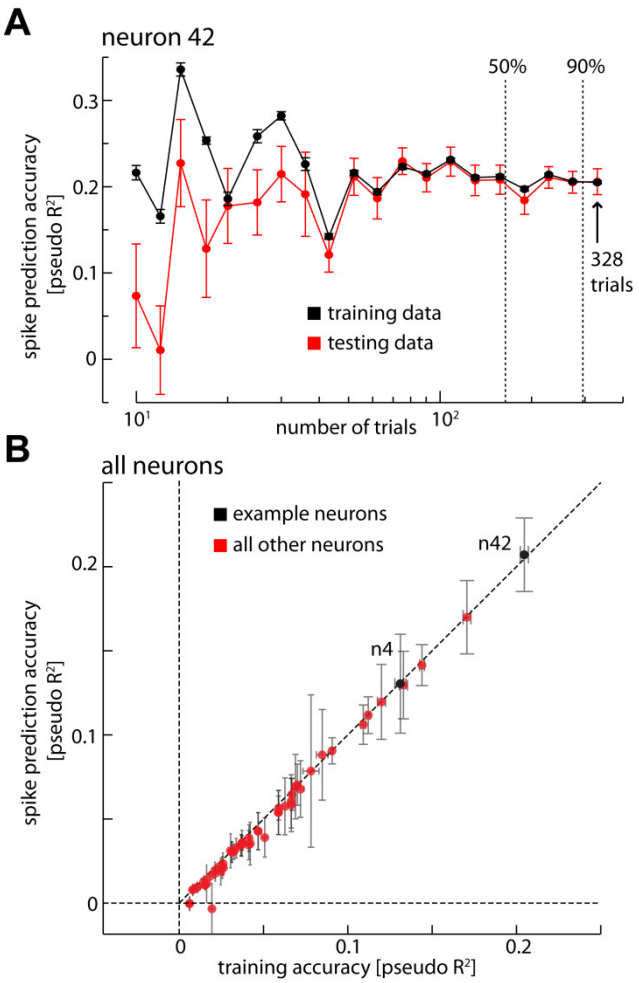


Figure 5.6. Model sensitivity and overfitting analysis. (*A*) Average of spike prediction accuracy (±SEM, 10-fold cross-validation) for the joint model on test data and on training data, as a function of the amount of data used. Total number of trials for this specific neuron is 328. Dashed vertical lines indicate thresholds for 50% and 90% for the trials. (*B*) Over-fitting analysis for the whole population. Error bars in both dimensions are ±2SEM.

The absence of improvement in spike prediction accuracy was not caused by the higher number of parameters in the joint model, since there is minimal overfitting (**Figure 5.6A** and **B**). Even though the full-saccade model explains away the saliency tuning modulation at fixation onset, it does not completely explain away saccade direction modulation at the end of the saccade (also see Supp. Information. and **Figure 5.12**). Furthermore, we checked that explaining away is robust within a considerable range of parameterizations of the temporal receptive fields (**Figure 5.13**). Thus, our finding that saliency is not represented in the FEF is not due to overfitting.

We observed that, for most of the neurons, saliency information alone allows some prediction of neural activity (**Figure 5.4C**). In fact, the spatio-temporal terms of the saliency model add predictive power (as measured by the pseudo R^2, $p<0.05$, bootstrap), to a model that considers only the purely temporal terms centered at fixation onset. However, the modulation related to saliency was explained away by including saccade information (**Figure 5.5A** and **B**). The fact that saliency related tuning is explained away seems surprising, since the relationship between saccades and saliency, although present, is fairly weak in our natural scene search task (**Figure 5.2C**). Even the apparently large effects in the saliency PSTHs (**Figure 5.4A** and **B**) and spike prediction (**Figure 5.4C** and **5.5A**) seem to be well explained based on these correlations (**Figure 5.5A-C**). Part of the directional tuning may be explained by the fact that the center of images tends to be more salient than the periphery (**Figure 5.2B**), and when fixation is at the edge of the image saccades toward the center become more likely. Furthermore, saccade onset and fixation onset happen close in time and saccade durations have some variability (**Figure 5.10B**). Neural responses are driven by a range of different factors. Ignoring some factors may lead us to draw wrong conclusions, but by modeling these factors together we can disambiguate which factors truly relate to the responses.

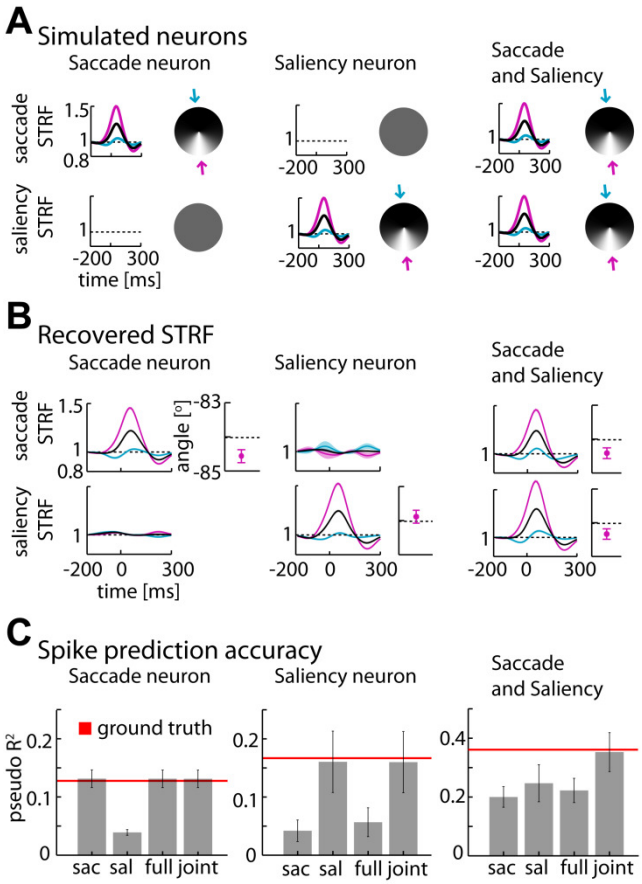


Figure 5.7. Simulations. (*A*) Simulated spatial and temporal filters for the three different kinds of neurons: purely saccade, purely saliency and both saccade and saliency encoding. We used a fixed temporal filter (triggered on saccade onset for saccade responses and on fixation onset for saliency responses) and a fixed preferred direction (represented by the circles with shades of grey – preferred direction corresponds to the lighter shades). Blue and purple curves correspond to the temporal gains in the directions of lower (blue arrow) and higher (purple arrow) modulation of the spatio-temporal receptive fields (STRFs). We simulated spikes using behavioral data corresponding to one neuron of our data set. (*B*) Recovered temporal filters (shaded area interval corresponds to ±2SEM) and preferred directions (±2SEM, 10-fold cross-validation. Black dashed line in error bar plot corresponds to the simulated/true preferred direction: direction of lighter shades of

grey signaled by the purple arrow in Panel A) for saccade and for saliency using the joint model for each of the simulated neurons of Panel A. (*C*) Cross-validated (± 2SEM, 10-fold) pseudo R^2 for each of the 4 models (saccade only, saliency only, full-saccade and joint model) for each of the 3 simulated neurons (saccade only, saliency only and both saccade and saliency).

Simulations

It could be that we failed to find true saliency responses in FEF because our data analysis routines did not correctly handle the correlations between the variables. We thus simulated equivalent amounts of data using a range of models: a purely saccade neuron, a purely saliency neuron and a neuron that encodes saccade and saliency simultaneously (**Figure 5.7A**, see Methods for details). We then asked if our methods would be able to recover the spatio-temporal tuning of these simulated neurons. Using the same GLM approach as above, we find that we can readily detect tuning to preferred saccade direction or saliency direction (**Figure 5.7B** and **C**) and the spatially and non-spatially dependent temporal responses. If the neurons in the actual FEF sample were truly tuned to the definition of saliency we are using, then these simulations demonstrate that we should have been able to reconstruct this dependence.

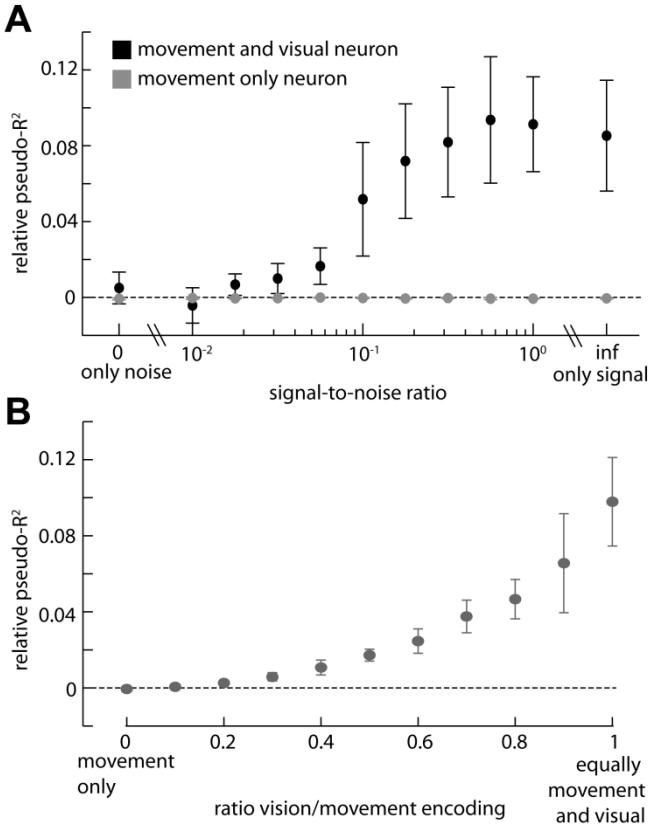


Figure 5.8. Statistical power analysis. (*A*) Relative pseudo R^2 between the joint model and the movement model, (±2SEM) as a function of the signal-to-noise ratio of the saliency definition, for a saccade only neuron and for a neuron that encodes saccade and saliency. (*B*) Relative pseudo R^2 between the joint model and the movement model (±2SEM), as a function of the amount of saliency that the neuron encodes relative to movement.

Lastly, it might simply be that our analysis was underpowered and more data would have been necessary to observe modulations in firing rate due to saliency. To test for this possibility we simulated neurons using the STRF component of the fitted receptive field to data of a particular neuron (neuron 4, **Figure 5.4B**; see Methods for details). We degraded the signal quality in our simulated neurons in two ways: (1) We made the definition of saliency used in the models worse by adding noise to the IK-saliency

definition and (2) We simulated neurons that were mostly tuned to saccade movement with progressively weaker modulation due to saliency. We found that even if saliency signals were highly corrupted (SNR~0.1) the amount of data available here should have been sufficient to resolve saliency related tuning (**Figure 5.8A**). We also found that even if the saliency tuning is substantially smaller than saccade tuning (by a factor of ~3) these effects should have been picked up (**Figure 5.8B**). Concretely, we can say that if IK-saliency had at least 25% influence on the neural activity of this neuron then we should have had more than 95% probability of finding it.

5.5 Discussion

In this study, we examined the activity of frontal eye field neurons to determine whether or not they represent bottom-up saliency while a monkey searches for small targets embedded in natural scenes. We found that saliency is mildly predictive of eye-movement direction during natural scene search but it appears not to be a determinant of FEF activity when other correlated, saccade-related covariates are properly taken into account. Our finding that FEF does not appear to represent bottom-up saliency suggests that the activity of the FEF may be dominated by top-down target-selection and saccade planning.

Our study has used eye movements during natural scene viewing to ask if neurons in the FEF represent bottom-up saliency. There are, of course, factors that limit the interpretation of our results.

Caveat 1: Our definition of saliency may differ from the actual representation of bottom-up saliency used by the FEF. We have employed a commonly used definition of bottom-up saliency (Itti and Koch, 2000). Past research has shown that most definitions of bottom-up saliency lead to saliency maps that are highly correlated with one another and are often difficult to disambiguate behaviorally (Borji et al., 2012). This is because

most computational definitions of bottom-up saliency effectively ask how dissimilar image patches are from the rest of the image and the specific metric of similarity often has little influence in such cases (Schölkopf and Smola, 2002). Therefore, it seems unlikely that other definitions of bottom-up saliency would have improved our ability to observe saliency tuning in FEF neurons.

Caveat 2: Our results show that if bottom-up saliency is represented in the FEF during natural scene search it is only explaining a tiny proportion of the overall activity. This does not imply that there is no representation of bottom-up saliency, nor does it imply that this proportion would be as small if it was a free-viewing task; just that our results support a weak representation. However, given that activity in the FEF is sufficiently strongly dominated by planning, it appears that bottom-up saliency representation is not a central function of FEF.

Previous research using artificial stimuli has suggested that significant activity in the FEF is devoted to the representation of visual saliency, noting that salient objects within the receptive field of an FEF cell may elicit high activity even without a saccade that actually ends in the receptive field (Bichot and Schall, 1999; Murthy et al., 2001; Thompson et al., 1997 ; Thompson et al., 1996). However, our results suggest that bottom-up saliency is not represented in the FEF. Furthermore, other studies using natural scenes suggest that visual cells do not respond to stimuli unless their receptive field contains the target of a future saccade (Burman and Segraves, 1994; Phillips and Segraves, 2010). How can this difference be explained? We suggest a couple of possible explanations for this apparent contradiction.

First, the eye-movement field has had some difficulty to adhere to a uniform definition of saliency, and generally includes a combination of bottom-up and top-down — including target relevance and the probability of a saccade — factors within the realm of saliency (but see Melloni et al.,

2012). This ambiguity makes it difficult to directly relate bottom-up saliency to activity in the FEF.

Second, we expect activity to be much higher for non-targets in a search task where the number of distractors is small (see McPeek and Keller, 2002). Given the small number of targets and the exceptionally high levels of saliency used in typical experiments, results may not generalize to search in natural scenes. Furthermore, it may be that highly salient stimuli trigger implicit planning of saccades that is later aborted, and hence, that the activity of a visual cell represents the amount of covert attention allocated to that location. Future work should directly compare the responses of FEF neurons to the traditional artificial salient stimuli and to more natural stimuli.

There are many computational definitions of the top-down factors that are likely to be represented in the FEF. The oculomotor system takes into account what the task-relevant target looks like (the relevance) (Serre et al., 2007) and the likely locations of the target given the scene context (the gist) (Torralba et al., 2006; Vogel and Schiele, 2007). Several studies have shown that most of search is driven by task-demands (Yarbus, 1967) and that it can override sensory-driven (bottom-up) saliency almost entirely (Einhäuser et al., 2008). In our task the monkey was not free-viewing but searching for an embedded target. Looking for representations of these top-down influences is possible with the methods presented here and would be an exciting topic for future research.

If bottom-up saliency is not represented in the FEF but it is important for the selection of saccades, it should be represented somewhere else. A model of a processing stream for visual saliency suggests a succession of stages in the visual-motor pathway from V1 to extrastriate visual cortex and on to areas LIP and FEF (Soltani and Koch, 2010). A recent imaging study has suggested that V1 represents bottom-up saliency while FEF is involved with target enhancement (Melloni et al., 2012). There have been reports supporting the existence of visual saliency maps in V4 (Burrows and Moore,

2009; Mazer and Gallant, 2003; Zhou and Desimone, 2011), LIP (Arcizet et al., 2011; Constantinidis and Steinmetz, 2005; Gottlieb et al., 1998), and FEF (Schall and Thompson, 1999; Thompson and Bichot, 2005; Wardak et al., 2010). A true bottom-up saliency map must represent the conspicuity of stimuli in the visual field, independent of the individual stimulus features themselves. However, given our results about the subtle ways by which apparent saliency tuning may arise, it seems fair to state that the question of if and where the brain represents saliency has not yet received a sufficient answer. It is not clear where in the visuomotor system relevance/target-matching is computed, but this study provides a counter-point to the hyper-salient tasks used in artificial experiments.

The approach taken here provides a template for how multiple factors that simultaneously might affect neural responses can be analyzed. Specifically, our analysis attempts to define what it means to say that the FEF encodes saliency when other correlated variables, such as saccade planning, may also be encoded by the same neurons. Here we used a precise definition of bottom-up saliency from the computational literature to quantify the extent to which FEF neurons represent bottom-up visual saliency during natural scene search. We found that it is not strongly represented. Instead, saccade planning and execution dominate the neural responses. This emphasizes the role of the FEF as a premotor structure, where neural activity encodes information about the importance of various spatial locations as potential saccade targets, independent of the visual properties of those locations.

5.6 Supplemental Information

Description of IK-Saliency algorithm and generalized linear model

We considered IK-saliency (Itti L and C Koch 2000; Walther D and C Koch 2006) as our definition of saliency. This is a bottom-up definition of saliency, i.e. based only on basic image features, independent of task objectives (see **Figure 5.9A**): (a) A total of 7 vision features (color channels tuned to red/green and blue/yellow hues, four orientations and brightness) are computed; (b) Each is computed at several different spatial scales using Gaussian pyramids as linear filters which consist of progressively low-pass filtering and subsampling; (c) This is followed by center surround differences across spatial scales, which compute local spatial contrast in each feature generating 6 maps for each feature - a total of 42 maps; (d) Non-linear iterative lateral inhibition incorporates center surround competition within each map. This iterative scheme uses Differences-of-Gaussians followed by a negative shift and half-wave rectification in order to suppress areas that are balanced in terms “excitation” and “inhibition” (with values near zero after the Differences-if-Gaussians is applied) and set every pixel to a non-negative value; (e) After competition, feature maps are combined into a single conspicuity map for each of the 3 feature types (color, intensity and orientation) and step d) - center surround competition - is repeated for each of the three conspicuity maps; (f) The three conspicuity maps are finally summed into the single map, the saliency map. We used the publicly available toolbox (http://www.saliencytoolbox.net/index.html (Walther D and C Koch 2006)) for computing IK-saliency for each image with the default parameter values and considered the three, equally-weighted, channels: color, intensity and orientation.

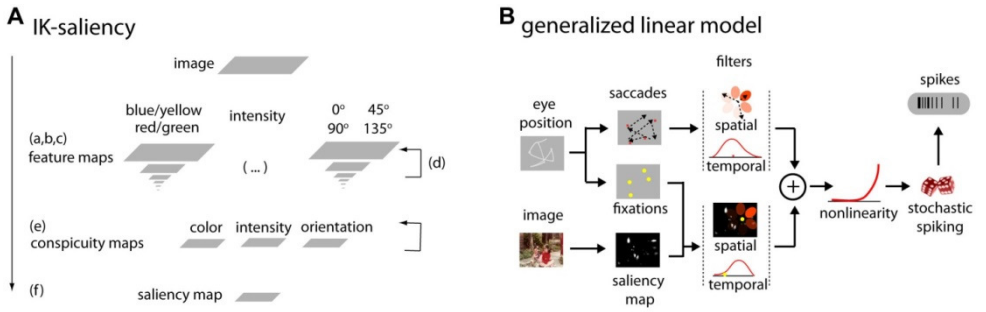


Figure 5.9. IK-saliency and generative model. (A) IK-saliency algorith. Scheme of the several steps of the the computation of IK-saliency (see description in the text above). (B) Generative model. We assume that the firing rate of the neuron depends on eye movement (saccades) and/or on the saliency of the image surrounding the fixation point. Each neuron has a preferred direction for saccade and a preferred direction for saliency. Also, each neuron has a temporal reaction to saccades and another one for saliencies that are centered on saccade onset and fixation onset respectively. We assume that the spiking activity is Poisson generated from the firing rate.

Saccade statistics and saccade modulation

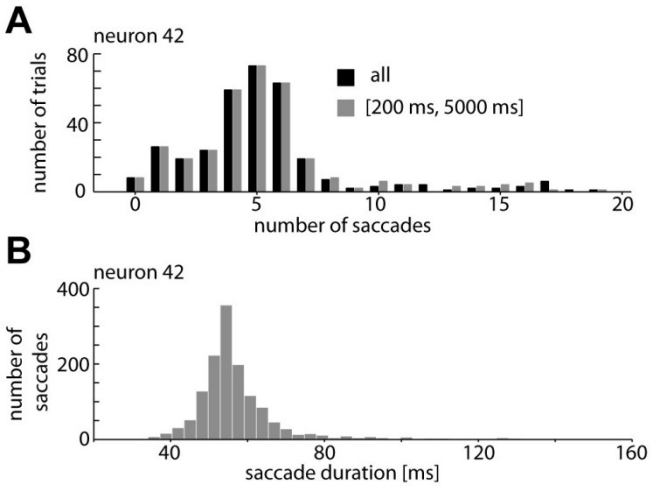


Figure 5.10. Saccade statistics. (A) Distribution of the number of saccades for the trials of one of the example neurons (neuron 42) using all the data (black) or only

the saccades that we considered (grey; the ones in the interval [200 ms, 5000 ms]). (B) Variability in saccade duration. Histogram of saccade durations for the eye movements during the trials for neuron 42. This variability could be the reason why, even for neurons only encode saccade, the saccade only model fails to completely explain away the joint model (see **Figure 5.5A** and **B**, center panels and **Figure 5.12**).

Saccade PSTH for example neuron number 42

For completeness we show the PSTH centered at saccade onset for neuron 42 (see **Figure 5.11**), the example neuron in **Figure 5.4A, B** and **5.5C**.

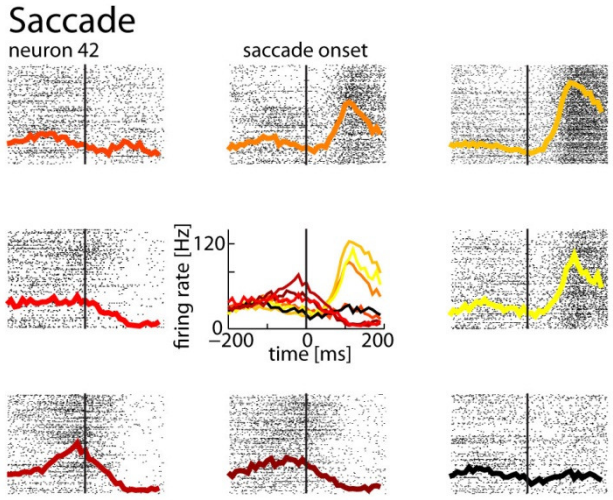


Figure 5.11. Saccade PSTH for neuron 42. Rasters sorted by direction of saccade, centered on saccade onset and the correspondent peri-stimulus time histograms (PSTHs) for neuron 42 from **Figure 5.4A, B** and **Fig. 5.5C**.

Full-saccade model doesn't explain away saccade spatial modulation at end of saccade

We tested an extra model - the complete saccade model- similar to the joint model but with saccade direction modulation at end of saccade/beginning of fixation instead of saliency modulation. We then computed the relative pseudo R2 of the complete saccade model relative to the full saccade model and observed that the full saccade model does not completely explain away the direction of saccade modulation at end of saccade/beginning of fixation. In other words, saccade direction information at end of saccade adds predictive power to the model. This suggests that saccade duration variability (see **Figure 5.10B**) is what prevents the saccade model from completely explaining away the joint model.

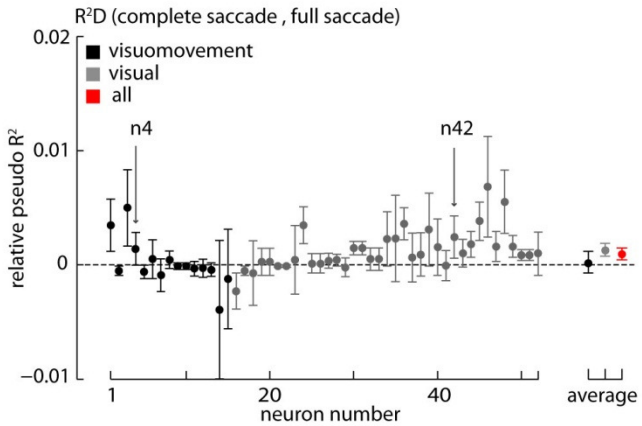


Figure 5.12. Predictive power of end of saccade modulation terms. Relative pseudo R2 between the complete saccade model and the full saccade model (±2SEM, 10-fold cross-validation for each individual neuron; 95% bootstrap confidence intervals for the averages across the recorded population and subpopulations). Arrows signal neurons 4 and 42, the example neurons of **Figure 5.3** and **5.4** respectively.

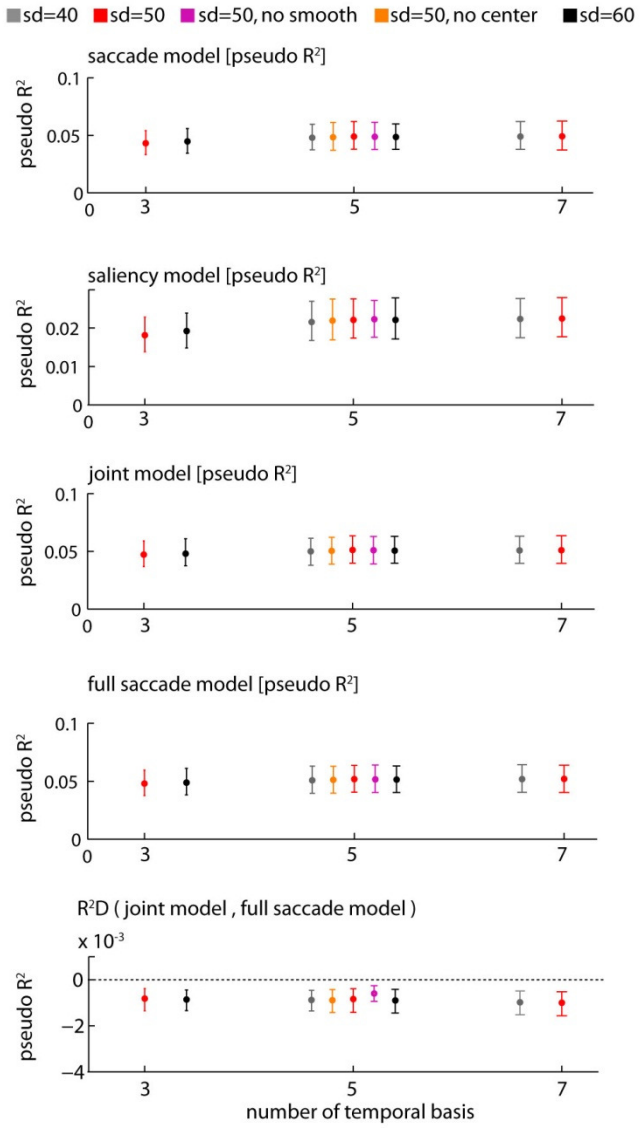


Figure 5.13. Robusteness of expaining away to change in some of the parameters. Pseudo R2 for the saccade, saliency, joint and full saccade models and relative pseudo R2 between the joint model and the full saccade model (95% confidence intervals, bootstrap across neurons). We considered not smoothing or not centering the IK-saliency definition and also several values for the parametrization of the temporal receptive fields, namely the number and standard deviation of the temporal basis functions.

Robustness of explaining away to changes in the parameters

We tested the robustness of our results to changes in the parameters of the models, specifically to width and number of basis functions that parameterize the temporal responses. We also tested whether centering (subtracting the mean) or smoothing (low-pass filtering) the IK-saliency definition had any influence in the conclusions of our analysis. We find that our results are robust to all these changes (see **Figure 5.13**).

6. Final Discussion

In this dissertation we have explored several approaches that aim at understanding decision making. The dissertation has essentially two parts; in a first part we use psychophysics and exclusively human behavioral data to characterize the generalization of prior expectations and to test prominent decision-making hypothesis (Chapters 2, 3 and 4). In a second part we use electrophysiology recordings to test the representation of relevant algorithmic variables in neural activity for making eye-movement decisions (Chapter 5).

6.1 Generalization and uncertainty

We started the dissertation by looking into generalization under uncertainty and of uncertainty itself. We did that by extending previous traditional motor-control generalization studies (Krakauer et al., 2000) to include uncertainty (Körding and Wolpert, 2004). In biology and in neuroscience in particular, the dream of finding physics-like general laws has always existed. Both generalization (Shepard, 1987) and Bayesian principles (Trommershäuser et al., 2011) have at a certain time been proposed as possible candidates for satisfying this utopian endeavor. It is thus exciting to bring them together in the research presented in Chapters 2 and 3.

We found that while uncertainty in the prior (the variance of a learned imposed perturbation) does not affect the generalization of the mean (Chapter 2), the mean does affects the generalization of uncertainty (Chapter 3). We show that, while also having a local component, uncertainty generalizes with a global component, and that manipulating the mean introduces an interesting asymmetry in the generalization of uncertainty

(Chapter 3). Furthermore we show that this asymmetry is consistent with the use of different reference frames when generalizing; target–centered for the mean and visual-feedback–centered for uncertainty.

While there are many theories about how the brain might represent or approximate probability distributions (Fiser et al., 2010; Hoyer and Hyvärinen, 2003; Ma et al., 2006), to our knowledge none of these theories has explicit predictions for how these probability distributions are learned/generalized from previous relatively similar experiences. The differences and interactions that we present in Chapters 2 and 3 between the generalizations of mean and variance constrain and pose a challenge to future attempts at extending these theories to generalization.

Apart from theoretical modeling, future directions of research can include further behavioral experiments but also imaging and electrophysiology studies to understand the neural basis of the asymmetries reported here. Regarding behavior studies, if the generalization of uncertainty is visual-feedback-centered then we expect the asymmetry to disappear if the mean of the perturbation is introduced and increased progressively. We also expect the location of the peak of uncertainty to change if we try different magnitudes of mean perturbation; e.g. the generalization pattern of a perturbation with a 15 degree mean should have a peak closer to the learning direction than of a 30 or 45 degree perturbation.

How does the brain represent the uncertainty about various quantities in order to generalize about them? In an imaging study using an estimation paradigm similar to the one used in in Chapter 4 (Vilares et al., 2012), we found that likelihood and prior uncertainty activated non-overlapping brain regions. An analogous study could look into whether mean and variance of the prior have non-overlapping representations in the visuomotor paradigm used in Chapters 2 and 3 to provide some insight on the representation of generalization. Recently, a similar center-out reaching paradigm has been

tested in non-human primates (Dekleva et al., 2013) and promises to give some perspective into how populations of neurons represent and learn the variance of prior and likelihood.

6.2 Decision-making theories

In Chapter 4 we tested a commonly implicit assumption about the two-alternative forced choice (2AFC) paradigm; that the just-noticeable difference (JND) measures sensory uncertainty. We show that in doing this we simultaneously test decision-making theories; MAP vs. Sampling/Matching. While the results from our experimental paradigm favor the assumption that the brain uses MAP algorithm, the most important output of the work presented in this chapter is that it makes the assumption explicit and suggests a way of testing it that should be extended to other tasks.

Even though in the first part of this dissertation we talk about representation of the value of some quantity (the mean of the prior) along with its uncertainty (the variance of the prior), it is important to note that these representations themselves are the outcome of an inference problem. For instance it may be argued that the retina solves an inference problem given photons; studies have shown that the retina already has predictive encoding and that the receptive fields of retinal ganglion cells can change to improve predictive coding under new environmental statistics (Hosoya et al., 2005). Similarly, the lateral geniculate nucleus (LGN), the primary visual cortex (V1) and other early visual areas may also solve inference problems, given retinal input – for instance, some studies suggest that prior is implicitly embedded in the neural tuning of sensory neurons (Ganguli and Simoncelli, 2012) – and so on. Hence, we are not dealing with a representation, but with the outcome of hierarchies of inference. At each stage the input (the

likelihood), is combined with the prior at that stage and this new posterior is the input (or likelihood?) for the next inference. While it might seem reasonable, assuming that prior and likelihood are two separate entities represented in the brain might limit the conclusions we can get from data and even make us draw the wrong conclusions if that assumption is not valid.

Following this line of thought there is another possible caveat in this study. It could be that we have changed the prior upstream of the input/likelihood whose uncertainty the JND is measuring; if the 2AFC measures likelihood uncertainty, and if we change the prior (uncertainty) upstream of this likelihood, then this JND should change with changes in prior variance. This is because this likelihood/input is in fact a posterior/output of a previous inference process that involves the prior that we are manipulating. Hence the JND would be actually measuring a posterior uncertainty – which changes with the variance of the prior even for the MAP hypothesis.

However, the fMRI study we mentioned above (Vilares et al., 2012) that uses a similar estimation task to the one we use here suggests that, at least for this task, the representations of prior and likelihood occur in different regions in the brain; while changes in likelihood uncertainty differentially activates earlier visual areas, changes in prior uncertainty differentially activate the putamen, amygdala, insula and orbitofrontal cortex. Hence, and even though the above mentioned hierarchical stream of inference most likely has recursive connections (Kok et al., 2013), we could argue that it is not very unreasonable to expect that, for our task, changes in prior are occurring mostly downstream of the likelihood of interest.

6.3 Neural representation of relevant variables

In Chapters 2 and 3 we explored the generalization of visuomotor rotations using a center-out reaching paradigm. Our results are compatible with a visual-feedback-centered reference frame for generalization of uncertainty. The importance of visual components in traditional center-out reaching paradigms has been highlighted before. For example, a visual aspect of targets, the saliency, has been shown to influence reaches (Wood et al., 2011). Saliency has also been shown to be a relevant variable for deciding where to look next (Berg et al., 2009; Einhäuser et al., 2006; Foulsham et al., 2011). In the second part of the thesis we looked into whether individual neurons of the monkey's frontal eye field (FEF) represent bottom-up saliency in a natural scene searching task. We found that even though saliency predicts eye movement, its predictive power gets explained away when we take into account saccade-related covariates.

Future research could investigate if the FEF represents other, non-bottom up, features that have been shown to predict fixation locations during natural scene, such as target related features (Ramkumar et al., 2013; Serre et al., 2007) as well as the likely locations of the target given the scene context, the gist (Torralba et al., 2006; Vogel and Schiele, 2007). Furthermore, future studies should also verify if the apparent absence of representation of bottom-up saliency still holds during free-viewing of natural scenes.

6.4 Concluding Remarks

Common to the different chapters of this thesis is the attempt to bridge or connect paradigms or theories from different sub-fields; in Chapter 2 and 3 we extend traditional visuomotor generalization studies in motor control to incorporate uncertainty by adapting a paradigm that has been used in several Bayesian-brain studies; In Chapter 4 we demonstrate the

connection between assumptions of the 2AFC paradigm and (Bayesian) decision-making theories. In Chapter 5 we used natural stimuli and an objective computational definition for testing the neural representation of bottom-up saliency, a relevant variable for making eye-movement decisions.

In conclusion, this dissertation gives several contributions to decision-making research at different levels; characterizes general principles in the generalization of probability distributions, tests which decision-making algorithms the brain might use and, finally, tests whether and how individual neurons represent specific variables relevant for deciding where to look next.

7. References

Ahrens, M.B., Paninski, L., and Sahani, M. (2008). Inferring input nonlinearities in neural encoding models. Network: Computation in Neural Systems *19*, 35-67.

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

Arcizet, F., Mirpour, K., and Bisley, J.W. (2011). A Pure Salience Response in Posterior Parietal Cortex. Cereb Cortex.

Barlow, H., Fitzhugh, R., and Kuffler, S. (1957). Change of organization in the receptive fields of the cat's retina during dark adaptation. The Journal of physiology *137*, 338-354.

Battaglia, P.W., Kersten, D., and Schrater, P.R. (2011). How haptic size sensations improve distance perception. PLoS computational biology *7*, e1002080.

Bell, A.J., and Sejnowski, T.J. (1997). The "independent components" of natural scenes are edge filters. Vision Res *37*, 3327-3338.

Bengio, Y., and Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. The Journal of Machine Learning Research *5*, 1089-1105.

Berg, D.J., Boehnke, S.E., Marino, R.A., Munoz, D.P., and Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. Journal of Vision *9*.

Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. Science *331*, 83.

Berniker, M., and Kording, K. (2008). Estimating the sources of motor errors for adaptation and generalization. Nature neuroscience *11*, 1454-1461.

Berniker, M., and Kording, K.P. (2011). Estimating the relevance of world disturbances to explain savings, interference and long-term motor adaptation effects. PLoS computational biology *7*, e1002210.

Berniker, M., Voss, M., and Kording, K. (2010). Learning priors for bayesian computations in the nervous system. PloS one *5*, e12686.

Bichot, N.P., and Schall, J.D. (1999). Effects of similarity and history on neural mechanisms of visual selection. Nature Neuroscience *2*, 549-554.

Bichot, N.P., Schall, J.D., and Thompson, K.G. (1996). Visual feature selectivity in frontal eye fields induced by experience in mature macaques. Nature *381*, 697-699.

Bizzi, E. (1968). Discharge of frontal eye field neurons during saccadic and following eye movements in unanesthetized monkeys. Exp Brain Res *6*, 69-80.

Bizzi, E., and Schiller, P.H. (1970). Single unit activity in the frontal eye fields of unanesthetized monkeys during head and eye movement. Exp Brain Res *10*, 151-158.

Borji, A., Sihite, D., and Itti, L. (2012). Quantitative Analysis of Human-Model Agreement in Visual Saliency Modeling: A Comparative Study.

Brashers-Krug, T., Shadmehr, R., and Bizzi, E. (1996). Consolidation in human motor memory. Nature *382*, 252-255.

Brayanov, J.B., Press, D.Z., and Smith, M.A. (2012). Motor Memory Is Encoded as a Gain-Field Combination of Intrinsic and Extrinsic Action Representations. The Journal of neuroscience *32*, 14951-14965.

Britten, K., Shadlen, M., Newsome, W., and Movshon, J. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. Journal of Neuroscience *12*, 4745-4765.

Bruce, C.J., and Goldberg, M.E. (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. Journal of neurophysiology *53*, 603.

Burge, J., Ernst, M.O., and Banks, M.S. (2008). The statistical determinants of adaptation rate in human reaching. Journal of Vision *8*, 20.21–19.

Burge, J., Fowlkes, C.C., and Banks, M.S. (2010). Natural-scene statistics predict how the figure–ground cue of convexity affects human depth perception. The Journal of neuroscience *30*, 7269-7280.

Burman, D.D., and Segraves, M.A. (1994). Primate frontal eye field activity during natural scanning eye movements. Journal of neurophysiology *71*, 1266-1271.

Burrows, B.E., and Moore, T. (2009). Influence and limitations of popout in the selection of salient visual stimuli by area V4 neurons. J Neurosci *29*, 15169-15177.

Cave, K.R., and Wolfe, J.M. (1990). Modeling the role of parallel processing in visual search. Cognit Psychol *22*, 225-271.

Christopoulos, V.N., and Schrater, P.R. (2009). Grasping objects with environmentally induced position uncertainty. PLoS computational biology *5*, e1000538.

Cisek, P., and Kalaska, J. (2002). Simultaneous encoding of multiple potential reach directions in dorsal premotor cortex. Journal of Neurophysiology *87*, 1149-1154.

Cisek, P., and Kalaska, J. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. Neuron *45*, 801-814.

Constantinidis, C., and Steinmetz, M.A. (2005). Posterior parietal cortex automatically encodes the location of salient stimuli. J Neurosci *25*, 233-238.

Crist, C.F., Yamasaki, D.S.G., Komatsu, H., and Wurtz, R.H. (1988). A grid system and a microsyringe for single cell recording. J Neurosci Methods *26*, 117-122.

Dekleva, B.M., Wanda, P.A., Kording, K.P., and Miller, L.E. (2013). The neural representation of likelihood uncertainty in the motor system. Paper presented at: Advances in Computational Motor Control (San Diego).

Deneve, S. (2008). Bayesian spiking neurons I: inference. Neural computation *20*, 91-117.

Dias, E.C., Kiesau, M., and Segraves, M.A. (1995). Acute activation and inactivation of macaque frontal eye field with GABA-related drugs. J Neurophysiol *74*, 2744-2748.

Dias, E.C., and Segraves, M.A. (1999). Muscimol-induced inactivation of monkey frontal eye field: effects on visually and memory-guided saccades. J Neurophysiol *81*, 2191-2214.

Diedrichsen, J., White, O., Newman, D., and Lally, N. (2010). Use-dependent and error-based learning of motor behaviors. The Journal of neuroscience *30*, 5159-5166.

DiMattina, C., Fox, S.A., and Lewicki, M.S. (2012). Detecting natural occlusion boundaries using local cues. Journal of vision *12*.

Donchin, O., Francis, J.T., and Shadmehr, R. (2003). Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: theory and experiments in human motor control. The Journal of neuroscience *23*, 9032-9045.

Doya, K. (2007). Bayesian brain: probabilistic approaches to neural coding (The MIT Press).

Duda, R.O., Hart, P.E., and Stork, D.G. (2012). Pattern classification (John Wiley & Sons).

Ehinger, K.A., Hidalgo-Sotelo, B., Torralba, A., and Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. Visual cognition *17*, 945-978.

Einhäuser, W., Kruse, W., Hoffmann, K.P., and König, P. (2006). Differences of monkey and human overt attention under natural conditions. Vision Research *46*, 1194-1209.

Einhäuser, W., Rutishauser, U., and Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. Journal of Vision *8*.

Elazary, L., and Itti, L. (2008). Interesting objects are visually salient. Journal of Vision *8*.

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

Everling, S., and Munoz, D.P. (2000). Neuronal correlates for preparatory set associated with pro-saccades and anti-saccades in the primate frontal eye field. J Neurosci *20*, 387-400.

Fecteau, J.H., and Munoz, D.P. (2006). Salience, relevance, and firing: a priority map for target selection. Trends Cogn Sci *10*, 382-390.

Fernandes, H.L., Albert, M.V., and Kording, K.P. (2011). Measuring generalization of visuomotor perturbations in wrist movements using mobile phones. PloS one *6*, e20290.

Fernandes, H.L., Stevenson, I.H., and Kording, K.P. (2012). Generalization of Stochastic Visuomotor Rotations. PloS one *7*, e43016.

Ferrier, D. (1875). Experiments on the brains of monkeys. Phil Trans, London (Croonian Lecture) *165*, 433-488.

Fetsch, C.R., Pouget, A., DeAngelis, G.C., and Angelaki, D.E. (2011). Neural correlates of reliability-based cue weighting during multisensory integration. Nature neuroscience *15*, 146-154.

Fiser, J., Berkes, P., Orbán, G., and Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. Trends in Cognitive Sciences.

Foulsham, T., Barton, J.J.S., Kingstone, A., Dewhurst, R., and Underwood, G. (2011). Modeling eye movements in visual agnosia with a saliency map approach: Bottom-up guidance or top-down strategy? Neural Networks.

Gaissmaier, W., and Schooler, L.J. (2008). The smart potential behind probability matching. Cognition *109*, 416-422.

Ganguli, D., and Simoncelli, E. (2012). Implicit embedding of prior probabilities in optimally efficient neural populations. arXiv preprint arXiv:12095006.

Geisler, W., Perry, J., Super, B., and Gallogly, D. (2001). Edge co-occurrence in natural images predicts contour grouping performance. Vision research *41*, 711-724.

Georgopoulos, A., Ashe, J., Smyrnis, N., and Taira, M. (1992). The motor cortex and the coding of force. Science *256*, 1692-1695.

Georgopoulos, A.P., Kalaska, J.F., Caminiti, R., and Massey, J.T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. The Journal of neuroscience *2*, 1527.

Ghahramani, Z., Wolpert, D.M., and Jordan, M.I. (1996). Generalization to local remappings of the visuomotor coordinate transformation. The Journal of neuroscience *16*, 7085-7096.

Girshick, A.R., Landy, M.S., and Simoncelli, E.P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nature neuroscience *14*, 926-932.

Gold, J.I., and Ding, L. (2013). How mechanisms of perceptual decision-making affect the psychometric function. Progress in neurobiology *103*, 98-114.

Goodbody, S., and Wolpert, D. (1998). Temporal and amplitude generalization in motor learning. Journal of Neurophysiology *79*, 1825-1838.

Gottlieb, J.P., Kusunoki, M., and Goldberg, M.E. (1998). The representation of visual salience in monkey parietal cortex. Nature *391*, 481-484.

Graham, K., Moore, K., Cabel, D., Gribble, P., Cisek, P., and Scott, S. (2003). Kinematics and kinetics of multijoint reaching in nonhuman primates. Journal of Neurophysiology *89*, 2667-2677.

Green, D.M., and Swets, J.A. (1966). Signal detection theory and psychophysics, Vol 1 (Wiley New York).

Harris, C.M., and Wolpert, D.M. (1998). Signal-dependent noise determines motor planning. Nature *394*, 780-784.

Harris, K.D., Csicsvari, J., Hirase, H., Dragoi, G., and Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. Nature *424*, 552-556.

Haslinger, R., Pipa, G., Lima, B., Singer, W., Brown, E.N., and Neuenschwander, S. (2012). Context Matters: The Illusive Simplicity of Macaque V1 Receptive Fields. PloS one *7*, e39699.

Hatsopoulos, N.G., Xu, Q., and Amit, Y. (2007). Encoding of movement fragments in the motor cortex. The Journal of neuroscience *27*, 5105.

Hays, A.V., Richmond, B.J., and Optican, L.M. (1982). A UNIX-based multiple process system for real-time data acquisition and control. Paper presented at: WESCON Conf Proc

Heinzl, H., and Mittlböck, M. (2003). Pseudo< i> R</i>-squared measures for Poisson regression models with over-or underdispersion. Computational statistics & data analysis *44*, 253-271.

Hinton, G.E., and Sejnowski, T.J. (1983). Optimal perceptual inference (IEEE Press Piscataway, NJ).

Hosoya, T., Baccus, S.A., and Meister, M. (2005). Dynamic predictive coding by the retina. Nature *436*, 71-77.

Howard, I.S., Ingram, J.N., Kording, K.P., and Wolpert, D.M. (2009). Statistics of natural movements are reflected in motor errors. J Neurophysiol *102*, 1902-1910.

Hoyer, P.O., and Hyvarinen, A. (2003). Interpreting neural response variability as Monte Carlo sampling of the posterior. Advances in neural information processing systems, 293-300.

Huang, V.S., Haith, A., Mazzoni, P., and Krakauer, J.W. (2011). Rethinking motor learning and savings in adaptation paradigms: model-free memory for successful actions combines with internal models. Neuron *70*, 787-801.

Hwang, E.J., Smith, M.A., and Shadmehr, R. (2006). Adaptation and generalization in acceleration-dependent force fields. Experimental brain research *169*, 496-506.

Hyvarinen, A., Hurri, J., and Vayrynen, J. (2003). Bubbles: a unifying framework for low-level statistical properties of natural image sequences. J Opt Soc Am A Opt Image Sci Vis *20*, 1237-1252.

Ingram, J.N., Howard, I.S., Flanagan, J.R., and Wolpert, D.M. (2010). Multiple grasp-specific representations of tool dynamics mediate skillful manipulation. Current Biology *20*, 618-623.

Ingram, J.N., Körding, K.P., Howard, I.S., and Wolpert, D.M. (2008). The statistics of natural hand movements. Exp Brain Res *188*, 223-236.

Itti, L., and Baldi, P. (2006). Bayesian surprise attracts human attention. Advances in neural information processing systems *18*, 547.

Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res *40*, 1489-1506.

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. Nature reviews neuroscience *2*, 194-203.

Jazayeri, M., and Shadlen, M.N. (2009). Probabilistic nature of time perception. Journal of Vision *9*, 1090.

Judge, S.J., Richmond, B.J., and Chu, F.C. (1980). Implantation of magnetic search coils for measurement of eye position: an improved method. Vision Res *20*, 535-538.

Kagerer, F.A., Contreras-Vidal, J., and Stelmach, G.E. (1997). Adaptation to gradual as compared with sudden visuo-motor distortions. Experimental brain research *115*, 557-561.

Kakei, S., Hoffman, D., and Strick, P. (1999). Muscle and movement representations in the primary motor cortex. Science *285*, 2136-2139.

Kayser, C., Kording, K.P., and Konig, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. Curr Opin Neurobiol *14*, 468-473.

Kayser, C., Nielsen, K.J., and Logothetis, N.K. (2006). Fixations in natural scenes: Interaction of image structure and image content. Vision Research *46*, 2535-2545.

Kayser, C., Salazar, R., and König, P. (2003). Responses to natural scenes in cat V1. J Neurophysiol *90*, 1910-1920.

Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol *4*, 219-227.

Kording, K. (2007). Decision theory: what "should" the nervous system do? Science *318*, 606-610.

Körding, K., Beierholm, U., Ma, W., Quartz, S., Tenenbaum, J., and Shams, L. (2007). Causal inference in multisensory perception. Plos One *2*.

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

Krakauer, J.W., Ghilardi, M.F., and Ghez, C. (1999). Independent learning of internal models for kinematic and dynamic control of reaching. Nature neuroscience *2*, 1026-1031.

Krakauer, J.W., Pine, Z.M., Ghilardi, M.F., and Ghez, C. (2000). Learning of visuomotor transformations for vectorial planning of reaching trajectories. J Neurosci *20*, 8916-8924.

Land, M.F., and Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? Vision Res *41*, 3559-3565.

Lewicki, M.S. (2002). Efficient coding of natural sounds. Nat Neurosci *5*, 356-363.

Liston, D.B., and Stone, L.S. (2008). Effects of prior information and reward on oculomotor and perceptual choices. The Journal of Neuroscience *28*, 13866-13875.

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

Ma, W.J. (2010). Signal detection theory, uncertainty, and Poisson-like population codes. Vision Research *50*, 2308-2319.

MacEvoy, S.P., Hanks, T.D., and Paradiso, M.A. (2008). Macaque V1 activity during natural vision: effects of natural scenes and saccades. J Neurophysiol *99*, 460-472.

Marr, D. (1982). Vision: A computational approach (Freeman & Co., San Francisco).

Mattar, A., and Ostry, D. (2007). Modifiability of generalization in dynamics learning. Journal of neurophysiology *98*, 3321-3329.

Mazer, J.A., and Gallant, J.L. (2003). Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. Neuron *40*, 1241-1250.

McPeek, R.M., and Keller, E.L. (2002). Saccade target selection in the superior colliculus during a visual search task. J Neurophysiol *88*, 2019-2034.

Melloni, L., van Leeuwen, S., Alink, A., and Müller, N.G. (2012). Interaction between Bottom-up Saliency and Top-down Control: How Saliency Maps Are Created in the Human Brain. Cerebral Cortex.

Mohler, C.W., Goldberg, M.E., and Wurtz, R.H. (1973). Visual receptive fields of frontal eye field neurons. Brain Res *61*, 385-389.

Möller, G., Einhäuser, W., and König, P. (2003). Cats' eye movements under natural conditions. Paper presented at: Soc Neurosci Abstr.

Moran, D., and Schwartz, A. (1999). Motor cortical representation of speed and direction during reaching. Journal of Neurophysiology *82*, 2676-2692.

Murthy, A., Thompson, K.G., and Schall, J.D. (2001). Dynamic dissociation of visual selection from saccade programming in frontal eye field. J Neurophysiol *86*, 2634-2637.

O'Shea, J., Muggleton, N.G., Cowey, A., and Walsh, V. (2004). Timing of target discrimination in human frontal eye fields. J Cogn Neurosci *16*, 1060-1067.

Oliva, A., Torralba, A., Castelhano, M.S., and Henderson, J.M. (2003). Top-down control of visual attention in object detection. 2003 International Conference on Image Processing, Vol 1, Proceedings, 253-256.

Olshausen, B., and Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature *381*, 607-609.

Palmer, J., Verghese, P., and Pavel, M. (2000). The psychophysics of visual search. Vision research *40*, 1227-1268.

Paz, R., Boraud, T., Natan, C., Bergman, H., and Vaadia, E. (2003). Preparatory activity in motor cortex reflects learning of local visuomotor skills. Nature neuroscience *6*, 882-890.

Paz, R., Nathan, C., Boraud, T., Bergman, H., and Vaadia, E. (2005). Acquisition and generalization of visuomotor transformations by nonhuman primates. Experimental Brain Research *161*, 209-219.

Pearl, J. (1988). Probabilistic reasoning in intelligent systems: networks of plausible inference (Morgan Kaufmann).

Pearson, T.S., Krakauer, J.W., and Mazzoni, P. (2010). Learning not to generalize: modular adaptation of visuomotor gain. Journal of neurophysiology *103*, 2938-2952.

Peng, X., Sereno, M.E., Silva, A.K., Lehky, S.R., and Sereno, A.B. (2008). Shape selectivity in primate frontal eye field. J Neurophysiol *100*, 796-814.

Phillips, A.N., and Segraves, M.A. (2010). Predictive activity in macaque frontal eye field neurons during natural scene searching. Journal of neurophysiology *103*, 1238-1252.

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E., and Simoncelli, E.P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature *454*, 995-999.

Poggio, T. (1990). A theory of how the brain might work. Cold Spring Harbor Symp Quant Biol *55*, 899-910.

Poggio, T., and Bizzi, E. (2004). Generalization in vision and motor control. Nature *431*, 768-774.

Pouget, A., and Snyder, L.H. (2000). Computational approaches to sensorimotor transformations. Nature neuroscience *3*, 1192-1198.

Ray, S., Pouget, P., and Schall, J.D. (2009). Functional distinction between visuomovement and movement neurons in macaque frontal eye field during saccade countermanding. J Neurophysiol *102*, 3091-3100.

Rickert, J., Riehle, A., Aertsen, A., Rotter, S., and Nawrot, M. (2009). Dynamic encoding of movement direction in motor cortical neurons. Journal of Neuroscience *29*, 13870-13882.

Robinson, D.A. (1963). A method of measuring eye movement using a scleral search coil in a magnetic field. IEEE Trans Bio-Med Electron *BME-10*, 137-145.

Rolls, E.T., and Tovee, M.J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. J Neurophysiol *73*, 713-726.

Roth, S., and Black, M.J. (2005). On the spatial statistics of optical flow. Paper presented at: Computer Vision, 2005 ICCV 2005 Tenth IEEE International Conference on (IEEE).

Sahani, M., and Dayan, P. (2003). Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity. Neural Computation *15*, 2255-2279.

Saijo, N., and Gomi, H. (2012). Effect of visuomotor-map uncertainty on visuomotor adaptation. Journal of neurophysiology *107*, 1576-1585.

Sakai, Y., and Fukai, T. (2008). When does reward maximization lead to matching law? PloS one *3*, e3795.

Saleh, M., Takahashi, K., Amit, Y., and Hatsopoulos, N.G. (2010). Encoding of Coordinated Grasp Trajectories in Primary Motor Cortex. The Journal of neuroscience *30*, 17079.

Sato, T., Murthy, A., Thompson, K.G., and Schall, J.D. (2001). Search efficiency but not response interference affects visual selection in frontal eye field. Neuron *30*, 583-591.

Sato, T.R., and Schall, J.D. (2003). Effects of stimulus-response compatibility on neural selection in frontal eye field. Neuron *38*, 637-648.

Sato, Y., and Aihara, K. (2011). A bayesian model of sensory adaptation. PloS one *6*, e19377.

Schall, J.D. (1991). Neuronal activity related to visually guided saccades in the Frontal Eye Fields of rhesus monkeys: Comparison with Supplementary Eye Fields. J Neurophysiol *66*, 559-579.

Schall, J.D., and Hanes, D.P. (1993). Neural basis of saccade target selection in frontal eye field during visual search. Nature *366*, 467-469.

Schall, J.D., Hanes, D.P., Thompson, K.G., and King, D.J. (1995). Saccade target selection in frontal eye field of macaque. I. Visual and premovement activation. J Neurosci *15*, 6905-6918.

Schall, J.D., and Thompson, K.G. (1999). Neural selection and control of visually guided eye movements. Annu Rev Neurosci *22*, 241-259.

Schiller, P.H., True, S.D., and Conway, J.L. (1980). Deficits in eye movements following frontal eye field and superior colliculus ablations. J Neurophysiol *44*, 1175-1189.

Schölkopf, B., and Smola, A.J. (2002). Learning with kernels : support vector machines, regularization, optimization, and beyond (Cambridge, Mass.: MIT Press).

Schwartz, O., and Simoncelli, E.P. (2001). Natural signal statistics and sensory gain control. Nat Neurosci *4*, 819-825.

Segraves, M.A. (1992). Activity of monkey frontal eye field neurons projecting to oculomotor regions of the pons. J Neurophysiol *68*, 1967-1985.

Segraves, M.A., and Goldberg, M.E. (1987). Functional properties of corticotectal neurons in the monkey's frontal eye field. J Neurophysiol *58*, 1387-1419.

Segraves, M.A., and Park, K. (1993). The relationship of monkey frontal eye field activity to saccade dynamics. J Neurophysiol *69*, 1880-1889.

Serences, J.T., and Yantis, S. (2006). Selective visual attention and perceptual coherence. Trends Cogn Sci *10*, 38-45.

Sergio, L., Hamel-Paquet, C., and Kalaska, J. (2005). Motor cortex neural correlates of output kinematics and kinetics during isometric-force and arm-reaching tasks. Journal of Neurophysiology *94*, 2353.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. IEEE Transactions on pattern analysis and machine intelligence, 411-426.

Shadmehr, R. (2004). Generalization as a behavioral window to the neural mechanisms of learning internal models. Human movement science *23*, 543-568.

Shadmehr, R., and Moussavi, Z. (2000). Spatial generalization from learning dynamics of reaching movements. Journal of Neuroscience *20*, 7807-7815.

Shadmehr, R., and Mussa-Ivaldi, F. (1994). Adaptive representation of dynamics during learning of a motor task. Journal of Neuroscience *14*, 3208-3224.

Sharpee, T., Rust, N.C., and Bialek, W. (2004). Analyzing neural responses to natural signals: maximally informative dimensions. Neural Comput *16*, 223-250.

Shea, C.H., and Kohl, R.M. (1990). Specificity and variability of practice. Research quarterly for exercise and sport *61*, 169-177.

Shepard, R.N. (1987). Toward a universal law of generalization for psychological science. Science *237*, 1317-1323.

Smith, E.C., and Lewicki, M.S. (2006). Efficient auditory coding. Nature *439*, 978-982.

Smyth, D., Willmore, B., Baker, G.E., Thompson, I.D., and Tolhurst, D.J. (2003). The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. J Neurosci *23*, 4746-4759.

Soltani, A., and Koch, C. (2010). Visual saliency computations: mechanisms, constraints, and the effect of feedback. J Neurosci *30*, 12831-12843.

Soltani, A., and Wang, X.J. (2009). Synaptic computation underlying probabilistic inference. Nature neuroscience.

Sommer, M.A., and Tehovnik, E.J. (1997). Reversible inactivation of macaque frontal eye field. Exp Brain Res *116*, 229-249.

Sommer, M.A., and Wurtz, R.H. (2000). Composition and topographic organization of signals sent from the frontal eye field to the superior colliculus. J Neurophysiol *83*, 1979-2001.

Sommer, M.A., and Wurtz, R.H. (2001). Frontal eye field sends delay activity related to movement, memory, and vision to the superior colliculus. J Neurophysiol *85*, 1673-1685.

Srivastava, A., Lee, A.B., Simoncelli, E.P., and Zhu, S.C. (2003). On Advances in Statistical Modeling of Natural Images. Journal of Mathematical Imaging and Vision *18*, 17-33.

Stocker, A.A., and Simoncelli, E.P. (2006). Noise characteristics and prior expectations in human visual speed perception. Nature neuroscience *9*, 578-585.

Suzuki, H., and Azuma, M. (1977). Prefrontal neuronal activity during gazing at a light spot in the monkey. Brain Res *126*, 497-508.

Tassinari, H., Hudson, T., and Landy, M. (2006). Combining priors and noisy visual cues in a rapid pointing task. Journal of Neuroscience *26*, 10154-10163.

Taylor, J.A., Hieber, L.L., and Ivry, R.B. (2012). Feedback-dependent Generalization. Journal of neurophysiology.

Taylor, J.A., and Ivry, R.B. Context-dependent Generalization. Frontiers in Human Neuroscience *7*, 171.

Taylor, J.A., and Ivry, R.B. (2011). Flexible cognitive strategies during motor learning. PLoS computational biology *7*, e1001096.

Theunissen, F.E., David, S.V., Singh, N.C., Hsu, A., Vinje, W.E., and Gallant, J.L. (2001). Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network *12*, 289-316.

Theunissen, F.E., Sen, K., and Doupe, A.J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci *20*, 2315-2331.

Thompson, K.G., and Bichot, N.P. (2005). A visual salience map in the primate frontal eye field. Progress in brain research *147*, 249-262.

Thompson, K.G., Bichot, N.P., and Sato, T.R. (2005). Frontal eye field activity before visual search errors reveals the integration of bottom-up and top-down salience. Journal of neurophysiology *93*, 337.

Thompson, K.G., Bichot, N.P., and Schall, J.D. (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. J Neurophysiol *77*, 1046-1050.

Thompson, K.G., Hanes, D.P., Bichot, N.P., and Schall, J.D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. J Neurophysiol *76*, 4040-4055.

Thoroughman, K., and Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. Nature *407*, 742-747.

Thoroughman, K., and Taylor, J. (2005). Rapid reshaping of human motor generalization. Journal of Neuroscience *25*, 8948-8953.

Torralba, A., Oliva, A., Castelhano, M.S., and Henderson, J.M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. Psychological review *113*, 766.

Treisman, A. (1988). Features and objects: the fourteenth Bartlett memorial lecture. Q J Exp Psychol A *40*, 201-237.

Trommershäuser, J., Körding, K.P., and Landy, M.S. (2011). Sensory cue integration (Oxford University Press).

Truccolo, W., Eden, U.T., Fellows, M.R., Donoghue, J.P., and Brown, E.N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. Journal of neurophysiology *93*, 1074.

Tseng, P.H., Carmi, R., Cameron, I.G.M., Munoz, D.P., and Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. Journal of Vision *9*.

Turano, K.A., Geruschat, D.R., and Baker, F.H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. Vision Res *43*, 333-346.

Turnham, E.J.A., Braun, D.A., and Wolpert, D.M. (2012). Facilitation of learning induced by both random and gradual visuomotor task variation. Journal of Neurophysiology *107*, 1111-1122.

van Beers, R.J. (2009). Motor learning is optimally tuned to the properties of motor noise. Neuron *63*, 406-417.

Van Hateren, J.H., and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. ProcRSocLond B Biol Sci *265*.

Verstynen, T., and Sabes, P.N. (2011). How each movement changes the next: an experimental and theoretical study of fast adaptive priors in reaching. The Journal of neuroscience *31*, 10050-10059.

Vilares, I., Howard, J.D., Fernandes, H.L., Gottfried, J.A., and Kording, K.P. (2012). Differential representations of prior and likelihood uncertainty in the human brain. Current Biology.

Vilares, I., and Kording, K. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. Annals of the New York Academy of Sciences *1224*, 22-39.

Vinje, W.E., and Gallant, J.L. (2002). Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. J Neurosci *22*, 2904-2915.

Vogel, J., and Schiele, B. (2007). Semantic modeling of natural scenes for content-based image retrieval. International Journal of Computer Vision *72*, 133-157.

Vul, E., Goodman, N.D., Griffiths, T.L., and Tenenbaum, J.B. (2009). One and done? Optimal decisions from very few samples. Paper presented at: Proceedings of the 31st annual conference of the cognitive science society.

Vulkan, N. (2000). An economist's perspective on probability matching. Journal of economic surveys *14*, 101-118.

Wainwright, M.J., Schwartz, O., and Simoncelli, E.P. (2002). Natural image statistics and divisive normalization: modeling nonlinearities and adaptation in cortical neurons. Statistical Theories of the Brain, 203–222.

Walther, D., and Koch, C. (2006). Modeling attention to salient proto-objects. Neural Networks *19*, 1395-1407.

Wardak, C., Vanduffel, W., and Orban, G.A. (2010). Searching for a salient target involves frontal regions. Cereb Cortex *20*, 2464-2477.

Wei, K., and Kording, K. (2009). Relevance of error: what drives motor adaptation? Journal of neurophysiology *101*, 655.

Wei, K., and Körding, K. (2010). Uncertainty of feedback and state estimation determines the speed of motor adaptation. Frontiers in computational neuroscience *4*, 11.

Weliky, M., Fiser, J., Hunt, R.H., and Wagner, D.N. (2003). Coding of natural scenes in primary visual cortex. Neuron *37*, 703-718.

Willmore, B., Watters, P.A., and Tolhurst, D.J. (2000). A comparison of natural-image-based models of simple-cell coding. Perception *29*, 1017-1040.

Wood, D.K., Gallivan, J.P., Chapman, C.S., Milne, J.L., Culham, J.C., and Goodale, M.A. (2011). Visual salience dominates early visuomotor competition in reaching behavior. Journal of vision 11.

Wozny, D.R., Beierholm, U.R., and Shams, L. (2010). Probability matching as a computational strategy used in perception. PLoS computational biology *6*, e1000871.

Wu, S., Chen, D., Niranjan, M., and Amari, S. (2003). Sequential Bayesian decoding with a population of neurons. Neural Computation *15*, 993-1012.

Yarbus, A.L. (1967). Eye movements and vision (New York: Plenum Press).

Zar, J.H. (1999). Biostatistical analysis. Prentice Hall New Jersey *4*.

Zemel, R., Dayan, P., and Pouget, A. (1998). Probabilistic interpretation of population codes. Neural computation *10*, 403-430.

Zhao, Q., and Koch, C. (2011). Learning a saliency map using fixated locations in natural scenes. Journal of Vision *11*.

Zhou, H., and Desimone, R. (2011). Feature-Based Attention in the Frontal Eye Field and Area V4 during Visual Search. Neuron *70*, 1205-1217.