Uncertainty is Maintained and Used in Working Memory

Aspen H. Yoo[1,2,3], Luigi Acerbi[1,2,4], Wei ji Ma[1,2]

Department of Psychology, New York University, NY, USA[1]

Center for Neural Science, New York University, NY, USA[2]

Department of Psychology, University of California, Berkeley, CA, USA[3]

Department of Computer Science, University of Helsinki, Helsinki, Finland[4]

Funding: This work was supported by NIH grant R01EY020958 to W.J.M. and training grant T32 EY7136-25 to A.H.Y.

1 Abstract

What are the contents of working memory? In both behavioral and neural computational models, the working memory representation of a stimulus is typically described by a single number, namely a point estimate of that stimulus. Here, we asked if people also maintain the uncertainty associated with a memory, and if people use this uncertainty in subsequent decisions. We collected data in a two-condition orientation change detection task; while both conditions measured whether people used memory uncertainty, only one required maintaining it. For each condition, we compared an optimal Bayesian observer model, in which the observer uses an accurate representation of uncertainty in their decision, to one in which the observer does not. We find that this "Use Uncertainty" model fits better for all participants in both conditions. In the first condition, this result suggests that people use uncertainty optimally in a working memory task when that uncertainty information is available at the time of decision, confirming earlier results. Critically, the results of the second condition suggest that this uncertainty information was maintained in working memory. We test model variants and find that our conclusions do not depend on our assumptions about the observer's encoding process, inference process, or decision rule. Our results provide evidence that people have uncertainty that reflects their memory precision at an item-specific level, maintain this information over a working memory delay, and use it implicitly in a way consistent with an optimal observer. These results challenge existing computational models of working memory to update their frameworks to represent uncertainty.

Keywords: visual working memory, Bayesian observer, optimal, uncertainty

2 Introduction

Visual working memory, the process involved in actively maintaining visual information over a short period, is essential for numerous everyday behaviors as "simple" as integrating visual information across saccades and as "complex" as reading comprehension, problem solving, and decision making (Baddeley & Hitch, 1974; Baddeley, 2003; Fukuda, Vogel, Mayr, & Awh, 2010; Conway, Kane, & Engle, 2003; Just & Carpenter, 1992). As important as it is, visual working memory is also a notoriously limited process, resulting in an imperfect and incomplete picture of the world it aims to represent.

Both behavioral (e.g., Zhang & Luck, 2008; Bays & Husain, 2008; van den Berg, Shin, Chou, George, & Ma, 2012; Fougnie, Suchow, & Alvarez, 2012) and neural (e.g., Ermentrout, 1998; Wang, 2001; Compte, 2006) models of visual working memory typically represent people's memory as a single number, a noisy estimate of the value of the stimulus. For example, someone may remember a 34° oriented line as 37°. It is, however, important in many visual working memory decisions to represent more than just a point estimate of the remembered stimulus, but the uncertainty as well. Uncertainty is technically defined as the width of a belief distribution over a stimulus, but intuitively represents how unsure an observer is about the stimulus. This is different from memory precision, which is how precisely an observer actually remembers the stimulus. An ideal observer's uncertainty will reflect the precision with which they remembered an item, such that they are less uncertain for more precise memories. In a variety of domains, taking uncertainty into account would increase performance and thus should be used. For example, high uncertainty over the memory of the location of a coffee cup may result in someone looking at it before reaching for it. High uncertainty over whether a friend changed their appearance may result in someone being less likely to comment on it.

Does uncertainty get taken into account in working memory-based decisions? An intuitive first place to look is the literature on working memory confidence, since confidence can be thought of as a readout of uncertainty. Experimenters have probed memory confidence by asking people to provide a rating (Rademaker, Tredway, & Tong, 2012; Vandenbroucke et al., 2014; Samaha & Postle, 2017), choose the best remembered item (Fougnie et al., 2012; Suchow, Fougnie, & Alvarez, 2017), or make a memory-based bet (Yoo, Klyszejko, Curtis, & Ma, 2018; Honig, Ma, & Fougnie, 2020). These studies have demonstrated that people have higher working memory confidence on trials that are remembered more accurately (but see Sahar, Sidi, & Makovski, 2020; Bona, Cattaneo, Vecchi, Soto, & Silvanto, 2013; Bona & Silvanto, 2014; Vlassova, Donkin, & Pearson, 2014; Maniscalco & Lau, 2015; Adam & Vogel, 2017; Samaha, Barrett, Sheldon, LaRocque, & Postle, 2016 for conflicting results), and a computational model in which memory judgements and confidence ratings are derived from the same underlying memory precision can quantitatively account for these joint data (van den Berg, Yoo, & Ma, 2017).

All these studies ask the participant to consciously access the quality of their memory. However, in naturalistic settings, people are typically not directly interrogated about their uncertainty, but use it implicitly in other decisions. For example, looking before reaching for one's coffee cup or commenting on a friend's appearance are decisions that presumably use uncertainty without conscious report. In this study, we take inspiration from perceptual decision-making studies, which have demonstrated that people implicitly incorporate uncertainty in a variety of decisions (e.g., van Beers, Sittig, & Gon, 1999; Ernst & Banks, 2002; Alais & Burr, 2004; Körding & Wolpert, 2004; Knill & Pouget, 2004; Ma, Navalpakkam, Beck, van den Berg, & Pouget, 2011; Jazayeri & Shadlen, 2010; Stocker & Simoncelli, 2006).

There is already some evidence that people use uncertainty implicitly in working memory-based decisions. Keshvari and colleagues had humans complete a four-item orientation change detection task (Keshvari, van den Berg, & Ma, 2012); Devkar and colleagues had humans and monkeys complete a three-item orientation change localization task (Devkar, Wright, & Ma, 2017). Stimuli in both studies were ellipses, which were independently assigned to be longer and narrower, providing "high-reliability" orientation information, or shorter and wider, providing "low-reliability" orientation information. The reliability of ellipses affected the precision with which they were encoded, and thus should have affected the memory uncertainty associated with each item. To maximize performance in both tasks, participants' uncertainty would need to reflect this variability in item-specific precision. Both studies found that a computational model that assumes participants use item-specific uncertainty accounted better for people's choices than alternative models.

Crucially, while these two studies provide evidence that people can implicitly use uncertainty, some experimental design choices do not allow us to conclude that people are actually maintaining uncertainty per se. First, participants in the study by Devkar and colleagues received trial-to-trial feedback on the correctness of their response. It is thus possible that participants simply learned a stimulus-response mapping (Maloney & Mamassian, 2009) rather than performing Bayesian inference or other forms of probabilistic computation (i.e., still using uncertainty in their decision; Ma, 2010). Second, precision in both studies was experimentally manipulated through ellipse reliability, which was held constant through a working memory delay. Thus, participants could have used this ellipse reliability as a proxy for uncertainty, rather than maintaining this exact information over the working memory delay (Barthelme & Mamassian, 2010).

Thus, the goal of this study was to investigate whether and how uncertainty is maintained over a working memory delay and used implicitly in a working memory task. To reach this goal, we collected data in a two-condition orientation change detection task and developed computational models to test different hypotheses about uncertainty. In the first condition, we established that people use uncertainty if a proxy to it is provided to them, replicating the results from Keshvari and others (2012). In the second condition, we asked if people still use uncertainty if this proxy is not provided at the time of decision. In other words, we asked if uncertainty is being maintained in working memory.

3 Experimental Methods

3.1 Participants

Thirteen participants (11 female; mean age $M = 21.1$ years, $SD = 2.5$) completed both conditions. All participants had normal or corrected-to-normal vision. Participants were naive to the study's hypotheses and were paid $12/hour and a $24 completion bonus. We obtained informed, written consent from all participants. The study was in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of New York University (IRB-FY2019-2490). Seven other participants were excluded because they did not meet performance criteria (explained in the Cross-session procedure section).

3.2 Stimuli

Stimuli were four, light-grey, oriented ellipses on a medium-grey background. Each ellipse could be long or short, to provide respectively higher or lower reliability information regarding the orientation of the ellipses. All ellipses had an area of 1.19 degrees of visual angle (dva). The high-reliability ellipse had an ellipse eccentricity of 0.9, such that the major axis and minor axes were 1.02 and 0.37 dva, respectively. The low-reliability ellipse eccentricity was determined separately for each participant to equate performance (details in Procedure).

On every trial, a stimulus display consisted of four ellipses. The probability of each ellipse being high reliability was 0.5, independent of the reliability of the other ellipses. The location of the first ellipse was drawn from a uniform distribution between polar angles 0° and 90°. Each ellipse after that was placed such that all ellipses were 90° apart on an imaginary annulus that was 7 dva away from fixation. Afterward, the x- and y- location of the ellipses were independently jittered -0.3 to 0.3 dva. In one condition, there were additionally oriented line stimuli, which were set to have approximately the same area as the ellipses. Stimuli were displayed on a 23 inch LED monitor with a refresh rate of 60 Hz and a resolution of 1920 x 1080 pixels.

3.3 Procedure

3.3.1 Trial Procedure

Ellipse condition. A trial began with a fixation cross presented for 1000 ms. Four ellipses were presented for 100 ms, followed by a 1000 ms delay, then by another four ellipses for 100 ms. On half of the trials, all ellipses in the second stimulus presentation were identical to the ellipses in the first stimulus presentation. On the other half of the trials, one ellipse changed in orientation. This change was drawn from a uniform distribution, so change of any magnitude had equal probability. Each ellipse had an equal probability of containing the change. “Change” and “no change” trials were randomly interleaved throughout the experiment. The participant indicated with a keyboard button press whether they believed there was an orientation change between the two displays.

Line condition. In the Line condition, the stimuli in the second presentation were oriented lines rather than ellipses. The task was otherwise identical. An example of a trial in the Ellipse and Line conditions is illustrated in Figure 1.

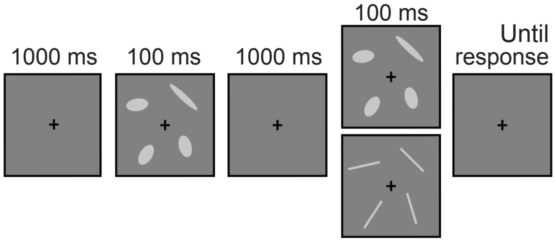

Figure 1: **Trial sequence.** Participants fixated on a cross, saw four ellipses (here showing one high-reliability ellipse and three low-reliability ellipses), maintained them over a delay, saw four stimuli again, and reported whether they believed there was an orientation change or not. In the Ellipse condition, ellipses in the second presentation were of the same reliability as in the first. In the Line condition, lines replaced ellipses in the second stimulus presentation, to avoid providing cues to the precision with which the first items were maintained.

3.3.2 Cross-Session Procedure

Participants completed both conditions over six one-hour sessions. They began their first session with a Practice block, designed to ease the participants into the task. They then completed 2000 trials of each condition, preceded by a Threshold block to set the "short" ellipse reliability for each condition. Participants completed all of one condition before completing the other, and the order was counterbalanced across participants. Participants were verbally informed that each trial had a 0.5 probability of a change occurring, that a change (if present) would occur in exactly one ellipse, and the change could be "of any magnitude; big changes are as possible as small changes." Participants were also verbally informed that some ellipses would be more elongated than others, that this may affect performance, and that half of the experiment would involve the stimuli changing from ellipses to lines. They were informed that their task did not change; the goal was always to indicate whether there was a change in orientation.

The Practice block consisted of 256 trials and was designed to ease naive participants into the speed of the task. The stimulus presentation time decreased throughout the course of the Practice block, from 333 ms to 100 ms, in 33 ms increments every 32 trials. Unlike the actual task, the ellipse eccentricities (i.e., reliabilities) of all ellipses within each trial were the same, but changed across trials. The stimuli in the second stimulus presentation corresponded to the condition that the participant completed first. For example, the stimuli in the second presentation were lines if the participant completed the Line condition first.

The Threshold block consisted of 400 trials and was used to set the ellipse eccentricity of the low-reliability ellipse in each condition. Like the Practice block, the ellipse eccentricities of all ellipses on each trial were the same, but changed on a trial-to-trial basis. The second stimulus presentation set were either ellipses or lines, corresponding to which condition the threshold was being set for. A cumulative normal psychometric function was fit to the accuracy as a function of ellipse eccentricity, and the low-reliability ellipse eccentricity was set as the value that corresponded to a predicted 65% accuracy. If the ceiling performance of the participant was estimated to be less than 75%, the Threshold block was repeated. If the psychometric function could not estimate an ellipse reliability for which performance would hit 65% after the second try, the participant was excluded from the experiment. Seven participants were excluded based on these criteria.

4 Experimental Results

The goal of our study was to investigate whether people maintained and used uncertainty implicitly in a working memory-based decision. To do this, we conducted a two-condition orientation change detection task. People could use memory uncertainty to maximize performance in both conditions, but only the Line condition required maintenance of that uncertainty. We conducted five repeated-measures ANOVAs to test whether condition (Ellipse, Line), the number of high-reliability ellipses displayed (N_{high}: 0, 1, 2, 3, 4), or their interaction significantly affected the following values: proportion report change, false alarm rate, hit rate (for all items), hit rate (when the changed item was a low-reliability ellipse), and hit rate (when the changed item was a high-reliability ellipse). These values are visualized in Figure 2, and the statistics are reported in Table 1.

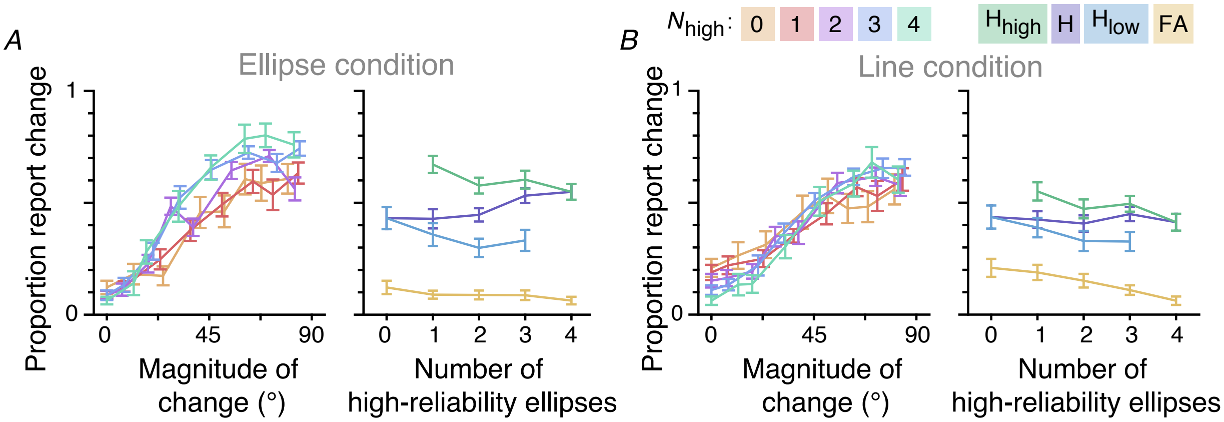

Figure 2: **Behavioral data.** Illustration of behavioral data for (*A*) Ellipse condition and (*B*) Line condition. For each condition, the left plots illustrates proportion report change as a function of magnitude of change. Data are binned by quantile, and different colored lines illustrate data from trials with different numbers of high-reliability ellipses presented on the first display. The right plots illustrate the proportion report change as a function of number of high-reliability ellipses, conditioned on whether there was no actual change (false alarm (FA): gold), a change in a low-reliability ellipse (H_{low}: blue), a change in a high-reliability ellipse (H_{high}: green), or a change in any ellipse (hit (H): purple). The color legend is displayed above the plots. Note that the aggregated hits are a weighted combination of the reliability-conditioned hits. The “Z” shape formed by the hit lines are an instance of Simpson’s paradox.

Dependent Variable	Factor	Statistics	p	ε	η^2
Proportion Report Change	N_{high}	$F(1.38,16.50)=3.37$	0.07	0.34	0.03
	Condition	$F(1,12)=1.33$	0.27	–	0.01
	N_{high} x Condition	**$F(2.12,25.38)=6.32$**	**0.005**	**0.52**	**0.04**
False Alarm Rate	N_{high}	**$F(1.93,23.17)=18.21$**	**2.07×10^{-5}**	**0.48**	**0.14**
	Condition	**$F(1,12)=6.50$**	**0.03**	–	**0.08**
	N_{high} x Condition	**$F(1.95,23.36)=4.94$**	**0.02**	**0.49**	**0.05**
Hit Rate (all)	N_{high}	**$F(1.36,16.30)=5.29$**	**0.03**	**0.34**	**0.04**
	Condition	$F(1,12)=2.47$	0.14	–	0.03
	N_{high} x Condition	**$F(2.04,24.48)=5.33$**	**0.01**	**0.51**	**0.03**
Hit Rate (low-reliability)	N_{high}	**$F(1.76,21.07)=23.26$**	**8.43×10^{-6}**	**0.59**	**0.08**
	Condition	$F(1,12)=0.29$	0.60	–	0.005
	N_{high} x Condition	$F(2.01,24.15)=0.37$	0.69	0.67	0.002
Hit Rate (high-reliability)	N_{high}	**$F(1.98,23.79)=35.44$**	**7.72×10^{-08}**	**0.66**	**0.13**
	Condition	**$F(1,12)=14.66$**	**0.002**	–	**0.17**
	N_{high} x Condition	$F(2.15,25.80)=0.75$	0.49	0.72	0.003

Table 1: **Results of two-way repeated-measures ANOVA**. Independent variables are N_{high} (0,1,2,3,4) and condition (Ellipse, Line), and dependent variables are displayed as the first column. Statistics of significant effects are bolded. For all ANOVAs, we report the Greenhouse-Geisser corrected results and ε (sphericity correction) when appropriate.

There was a statistically significant interaction between N_{high} and condition on proportion report change. In only the Ellipse condition, the proportion report change was modulated by the number of high-reliability ellipses (left plot of Fig. 2 A, B). There were significantly more false alarms in the Line condition ($M = 0.14$, $SEM = 0.03$) than in the Ellipse condition ($M = 0.09$, $SEM = 0.02$; purple lines in right plots of Fig. 2 A, B). Perhaps people mistook the change in stimulus as a change in orientation. Both reliability-conditioned hit rates (blue and green lines in right plots of Fig. 2 A, B) decreased with increasing N_{high}. Additionally, participants had significantly lower high-reliability hits in the Line condition and the Ellipse condition.

There is an interesting reverse in the qualitative trend when looking at all hit rates across all trials: hit rate increases as a function of N_{high}. This Simpson's paradox is a result of weighted averaging and the performance difference between the reliability-conditioned hit rates. As the number of high-reliability ellipses in a display increases, so does the probability of a change occurring in a high-reliability ellipse. Thus, the total hit rates for higher N_{high}s contain more high-reliability hits than low-reliability hits, driving this value upward. Similarly, the trials to compute hit rates for lower N_{high}s predominantly contain changes in low-reliability ellipses, thus driving the average downward. There was also a significant interaction between condition and N_{high} because this increase was greater in the Ellipse condition. This unsurprisingly reflects the results of proportion report change, for they are related values.

These statistics show that differences between factors and conditions exist, but are dissatisfying because they do not offer explanations of what these differences mean. In this paper, we take an approach to understand underlying working memory processes through computational modeling. Computational modeling plays a crucial role in the interpretation of data and the understanding of the underlying working memory and decision-making processes (Levenstein et al., 2020; Edwards, 1954). It allows us to make explicit assumptions and precise quantitative predictions, which provide committal, falsifiable explanations of the processes involved.

5 Modeling Methods

To test whether people are maintaining and using uncertainty when making their change detection decision, we use Bayesian observer models. Bayesian models provide a normative, flexible, and interpretable framework to study the working memory process. These models are particularly useful in cases where the observer is trying to make a decision without full knowledge of task-relevant information. In working memory, people do not have full knowledge because information is not remembered perfectly. While Bayesian decision theory describes how an observer should behave in order to maximize performance, different components of the model can be easily substituted with incorrect beliefs or suboptimal use of information, and thus provides a good template for building models with "imperfectly optimal observers" (Maloney & Zhang, 2010) or "model mismatch" (Orhan & Jacobs, 2014; Beck, Ma, Pitkow, Latham, & Pouget, 2012; Acerbi, Ma, & Vijayakumar, 2014).

We model the observer's decision process as consisting of an encoding stage and decision stage. The encoding stage describes the task statistics and our assumptions about how memories are generated. In the decision stage, the observer calculates a decision variable based on their belief of the encoding stage and decides whether to report Change or No change based on some decision rule. We compared two models: one in which uncertainty is maintained and used and another that is not, named the "Use Uncertainty" and the "Ignore Uncertainty" model, respectively. This section describes how these models were defined, fit, and compared.

5.1 Encoding Stage

In this section, we define the statistical structure of the experiment and define our assumptions about how memories are generated in an observer. On every trial, there is a 0.5 probability of there being a change, $p(C=1)=0.5$, where C takes values 0 (no change) and 1 (change). On change trials, exactly one item changes in its orientation, and each item is equally probable to be changed. The orientation change, Δ, is drawn from a uniform distribution, $p(\Delta)=\frac{1}{2\pi}$. (For mathematical convenience, and without loss of generality, we doubled the actual orientation of stimuli in all model specifications such that the values span 0 to 2π rather than 0 to π. We do not double these values when illustrating model fits.)

We denote the vector of all orientations of the items presented on the first display by $\boldsymbol{\xi}$, in which each element is an independent draw from a uniform distribution over orientation space. The vector of orientations at the second display, $\boldsymbol{\phi}$, was identical to $\boldsymbol{\xi}$ in no change trials. In change trials, the ith element of $\boldsymbol{\phi}$, the location of change, was equivalent to $\xi_i + \Delta$.

We model the memory process for each item of each display according to the Variable Precision model (van den Berg et al., 2012; Fougnie et al., 2012), by which memories are described as a continuous resource that randomly fluctuates across items and trials. The noisy measurements of each item on each display, $\boldsymbol{x} = (x_1, ..., x_N)$ and $\boldsymbol{y} = (y_1, ..., y_N)$, are conditionally independent and drawn from a Von Mises distribution centered on the actual orientation presentation,

$$p(\boldsymbol{x}|\boldsymbol{\xi};\boldsymbol{\kappa}_x) = \prod_{i=1}^{N} p(x_i|\xi_i,\kappa_{x,i}) = \prod_{i=1}^{N} \frac{1}{2\pi I_0(\kappa_{x,i})} e^{\kappa_{x,i}\cos(x_i-\xi_i)}$$
$$p(\boldsymbol{y}|\boldsymbol{\phi};\boldsymbol{\kappa}_y) = \prod_{i=1}^{N} p(y_i|\phi_i,\kappa_{y,i}) = \prod_{i=1}^{N} \frac{1}{2\pi I_0(\kappa_{y,i})} e^{\kappa_{y,i}\cos(y_i-\phi_i)}.$$

The κs are the concentration parameter of the Von Mises distribution, and are related to the precision with which each item is remembered; a higher κ corresponds to higher precision. The subscript of each κ indicates which item it refers to (e.g., $\kappa_{x,i}$ is concentration parameter for x_i, the i^{th} item the first stimulus presentation). We assume that memory precision varies across items, above and beyond the precision differences due to stimulus reliability. In other words, $\kappa_{x,i}$ and $\kappa_{y,i}$ are themselves random variables, rather than single values. Rather than sampling κ itself, we sample the Fisher information of the Von Mises distribution, J, from a gamma distribution:

$$p(J) = \frac{1}{\Gamma\left(\frac{\bar{J}}{\tau}\right)\tau^{\bar{J}/\tau}} J^{\frac{\bar{J}}{\tau}-1} e^{J/\tau},$$

where τ is the scale parameter of the gamma distribution and $\bar{J}$ is the mean precision. The relationship between J and κ is the following:

$$J = \kappa \frac{I_1(\kappa)}{I_0(\kappa)},$$

where I_0 is a modified Bessel function of the first kind of order 0 and I_1 is a modified Bessel function of the first kind of order 1 (van den Berg et al., 2012; Keshvari et al., 2012). We allow the mean precision to differ across stimulus shape; the precisions of memories corresponding to low-reliability ellipses are drawn from a gamma distribution with mean $\bar{J}_{\text{low}}$ and high-reliability ellipses with mean $\bar{J}_{\text{high}}$. Parameter τ is shared across both distributions. Because items in the first display were presented earlier, there are certainly differences in the precision with which items in the first and second display are maintained, independent of ellipse reliability. However, the amount that the first and second displays contribute to the overall measured change are extremely hard to tease apart in the model. Thus, we use one parameter per reliability and recognize that this estimate will be some average of the precisions of the first and second display.

When modeling the Line condition, we have an additional parameter, $\bar{J}_{\text{line}}$, which corresponds to the mean precision with which each line on the second display is remembered by the observer. To limit model complexity, the gamma function from which each line's precision is drawn shares the same scale parameter τ as the distributions from which the ellipse precisions are drawn.

5.2 Decoding Stage

5.2.1 Decision Variable

The essence of Bayesian inference is that an observer can compute a posterior over task-relevant latent variables, and should if they want to maximize performance. In this case, the observer should calculate the probability of the state of the world (i.e., change or no change) given their observations, $p(C|\boldsymbol{x},\boldsymbol{y})$, which they can compute using Bayes rule. With a scenario in which there are only two states of the world, it is convenient to combine these into a ratio. Thus, we assume the observer calculates, for each item, the ratio of the likelihood of there being change and the likelihood of there being no change:

$$d = \frac{p(C=1|\boldsymbol{x},\boldsymbol{y})}{p(C=0|\boldsymbol{x},\boldsymbol{y})} = \frac{p(\boldsymbol{x},\boldsymbol{y}|C=1)p(C=1)}{p(\boldsymbol{x},\boldsymbol{y}|C=0)p(C=0)}. \tag{1}$$

Details of the derivation can be found in Appendix 8.1, but this simplifies to the following expression:

$$d = \frac{p(C=1)}{p(C=0)} \frac{1}{N} \sum_{i=1}^{N} d_i, \tag{2}$$

where

$$d_i = \frac{I_0(\kappa_{x,i}) I_0(\kappa_{y,i})}{I_0\left(\sqrt{\kappa_{x,i}^2 + \kappa_{y,i}^2 + 2\kappa_{x,i}\kappa_{y,i}\cos(x_i - y_i)}\right)}. \tag{3}$$

I_0 is a modified Bessel function of the first kind of order 0, and the κs are the concentration parameters of the noise distributions for the item indicated in the subscript. Intuitively, d_i provides a measure of the evidence of change for the i^{th} item. It increases with the measured amount of change, $x_i - y_i$, weighted by a function of the precisions with which x_i and y_i are remembered. The d_is are averaged in the decision variable d, providing the optimal measure of evidence of change of the entire display.

This is the step in which the use of uncertainty comes in. Observers who correctly maintain and use uncertainty (i.e., observers who act in accordance with the optimal, "Use Uncertainty" model) compute exactly d_i as described. However, observers acting in accordance with the "Ignore Uncertainty" model do not know or do not consider that the precision of their memories for all items in both displays varies. Computing the decision rule for the Ignore Uncertainty observer is the same as replacing all κs in Eq. 3 with a constant, resulting in the following local decision variable:

$$d_i = \frac{I_0^2(\kappa_{\text{ass}}))}{I_0\left(\kappa_{\text{ass}}\sqrt{2 + 2\cos(x_i - y_i)}\right)}, \tag{4}$$

where κ_{ass} is the assumed precision for all items on all displays. The decision variable thus becomes just a function of $\cos(x_i - y_i)$, because the remainder of the expression is constant.

5.2.2 Decision Rule

The observer maps this decision variable onto a response by reporting Change whenever the probability of there being a change is greater than 0.5. An optimal observer would thus respond Change when the ratio of the likelihood of there being a change and the likelihood of there being no change (Eq. 2) is greater than 1 (Figure 3). However, we allow the observer to have some response bias (e.g., due to unequal priors, rewards, or motor costs), and thus implement the following decision rule:

$$\frac{1}{N}\sum_{i=1}^{N} d_i > k, \tag{5}$$

where k is a free parameter. For both models, we implemented global decision noise by adding zero-mean Gaussian noise with standard deviation σ_d to the log of decision variable d (Keshvari et al., 2012; Acerbi et al., 2014; Mueller & Weidemann, 2008). Additionally, participants randomly guess with probability λ, due to factors such as lapses in attention.

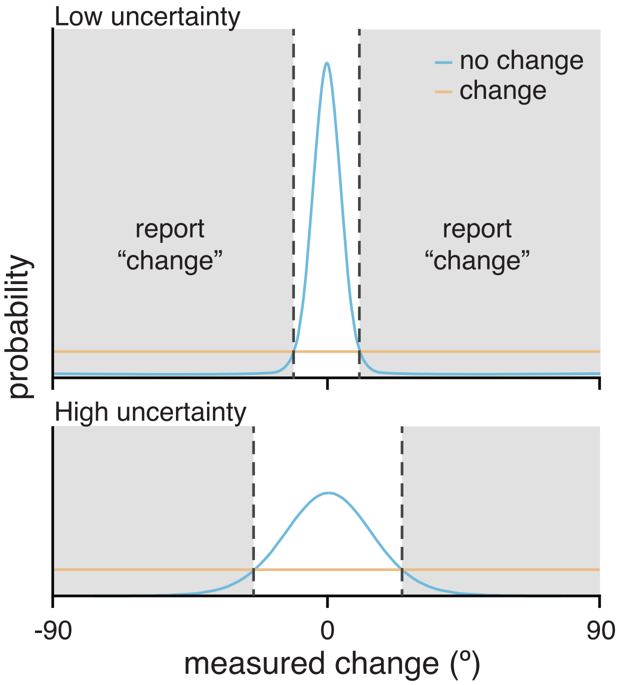

Figure 3: **Model didactics.** This didactic illustrates a simplified one-item version of this task. The probability of the measured change for an item given that the item did (orange) or did not (blue) change orientation, as estimated by the optimal observer. Uncertainty modulates the width of the no change distribution, such that higher uncertainty makes the no change distribution wider (bottom). The optimal observer (with $k = 0$) places their decision boundaries at the intersection of the change and no change distributions (vertical dashed lines), reporting "change" whenever that state of the world is more probable (shaded region) and "no change" otherwise.

5.3 Parameter Estimation and Model Comparison

5.3.1 Parameters

Both models in both conditions have parameters $\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, k, \lambda$, and σ_{d}. Parameters $\bar{J}_{\text{high}}$ and $\bar{J}_{\text{low}}$ correspond to the mean precision of the high- and low-reliability ellipses, respectively. Precision is also affected by the scale parameter of the gamma distribution from which item-wise precision is drawn, τ; this value is shared across the two ellipse types and the line when applicable. Parameter k is the observer's response bias; λ is the probability on each trial that the observer lapses and responds randomly; σ_{d} is the standard deviation of the Gaussian from which decision noise is simulated.

When fitting data from the Line condition, there is an additional parameter $\bar{J}_{\text{line}}$, corresponding to the mean precision with which the line stimulus is represented. The Ignore Uncertainty model has one additional parameter: J_{ass}, the assumed precision of all stimuli in both displays.

5.3.2 Parameter Estimation

The likelihood of the parameter combination $\boldsymbol{\theta}$ for a given participant and model is the probability of the data given the parameter combination. We used the log likelihood, which we denote LL:

$$\begin{aligned} \text{LL}(\boldsymbol{\theta}) &= \log p(\boldsymbol{\theta}|\text{data, model}) \\ &= \log \prod_{t}^{N_{\text{trials}}} p(r_t|\boldsymbol{\theta}) \\ &= \sum_{t}^{N_{\text{trials}}} \log p(r_t|\boldsymbol{\theta}), \end{aligned}$$

where r_t is the participant's response on the t^{th} trial. For each participant, we used maximum-likelihood estimation to find which parameter combination best describes participant's data. Computing the LL analytically is intractable, so we used Inverse Binomial Sampling (van Opheusden, Acerbi, & Ma, 2020), a method which efficiently computes an unbiased estimate of the LL. This calculation is stochastic, so we additionally used an optimization algorithm that can account for stochasticity and expensive LL evaluations (BADS; Acerbi & Ma, 2017). BADS explicitly incorporates uncertainty in the estimated LL and converges in fewer function evaluations than other stochastic optimization methods (e.g., CMA-ES, genetic algorithms), making it an ideal optimization method when likelihood calculations are computationally expensive and stochastic. We used 20 different starting positions, using Latin hypercube sampling, to reduce the probability of finding a local minimum. We took the parameter combination corresponding to the minimum negative log-likelihood of our runs as the ML parameter estimate. The estimated LL at the candidate optimum was reevaluated using 1000 samples in IBS, in order to reduce the standard deviation of estimation noise to less than 1. We denote the maximum log-likelihood by LL*.

5.3.3 Model Comparison

We compared models using corrected Akaike Information Criterion (AICc; Hurvich & Tsai, 1987) and the Bayesian Information Criterion (BIC; Schwarz, 1978). BIC penalizes for additional parameters harsher than AICc.

$$\mathrm{AICc} = -2\mathrm{LL}^* + 2N_{\mathrm{pars}} + \frac{2N_{\mathrm{pars}}(N_{\mathrm{pars}}+1)}{N_{\mathrm{trials}} - N_{\mathrm{pars}} - 1}$$

$$\mathrm{BIC} = -2\mathrm{LL}^* + 2N_{\mathrm{pars}} \log N_{\mathrm{trials}}$$

5.4 Modeling Results

We compared the fits of the Use Uncertainty and Ignore Uncertainty models to each of the conditions separately. The Use Uncertainty model provides a good qualitative fit to the data in both conditions (top row of Figure 4A), while the Ignore Uncertainty model is unable to capture the data (bottom row of Figure 4A). This result is reflected quantitatively (Figure 4B). For both conditions, all of the 13 participants were better fit by the Use Uncertainty model than the Ignore Uncertainty model (median [95% bootstrapped confidence interval (CI)] ΔAICc across subjects – Ellipse: 208 [109, 368], Line: 108 [58, 370]. ΔBIC – Ellipse: 221 [122, 381], Line: 122, [71, 383]). The summed ΔAICc and ΔBIC are consistent and are available in Appendix 8.4. Parameter estimates for the Use Uncertainty model in the Ellipse and Line condition can be found in the Appendix (Tables 4 and 5, respectively).

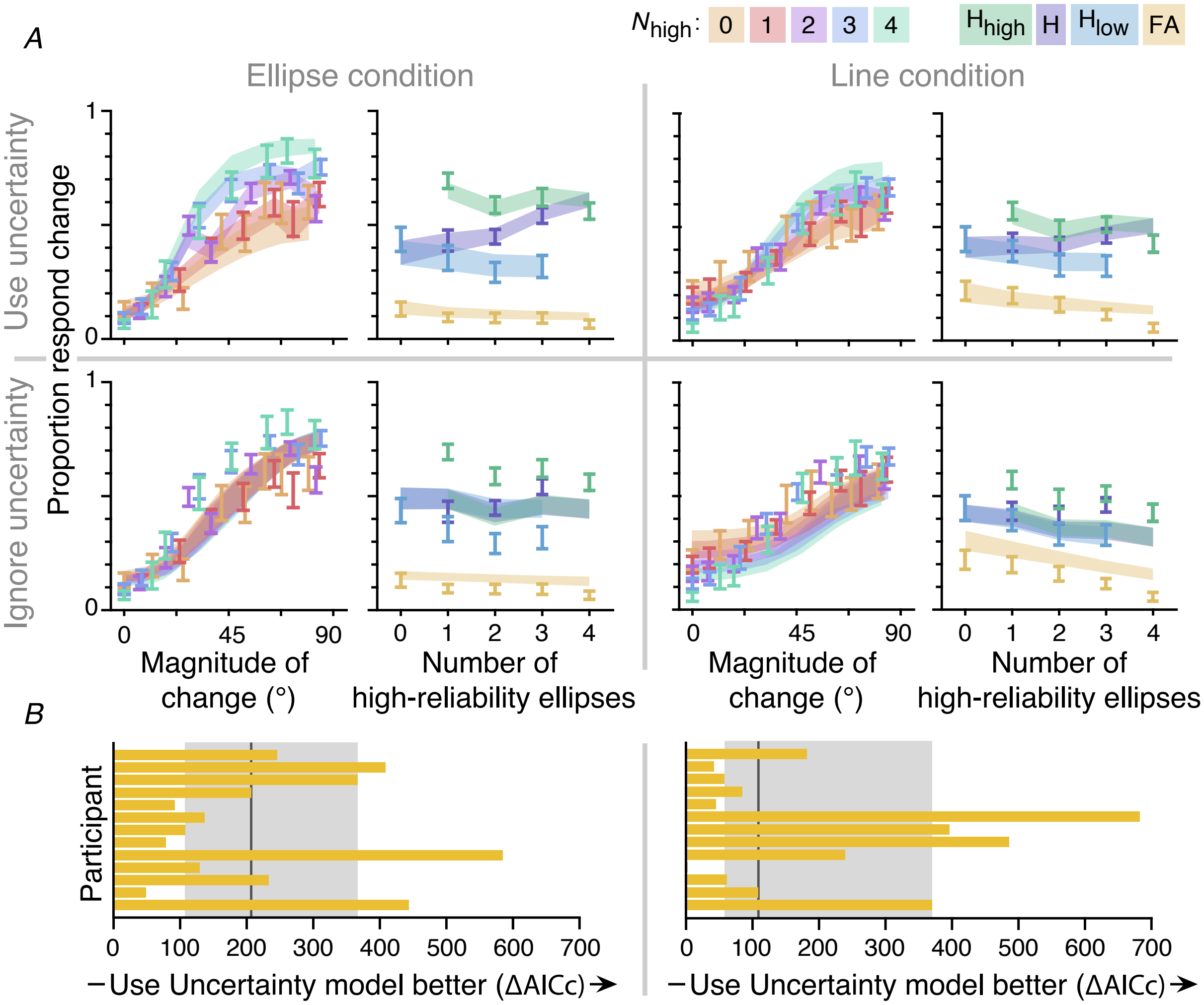

Figure 4: **Model fits.** *A*. $M \pm SEM$ data (error bars) and model fits (fills) for the Use (top) and Ignore (bottom) Uncertainty models and the Ellipse (left) and Line (right) conditions. For each model and condition, the left graph illustrates the proportion report change as a function of amount of change. Data and models are binned by quantiles, and color indicates the number of high-reliability ellipses. The right graph illustrates the proportion hits for high-reliability items (green), hits for low-reliability items (blue), total hits (purple), and false alarms (gold) as a function of number of high-reliability items. *B*. Model comparison for the Ellipse (left) and Line (right) conditions. Each bar indicates the individual-subject ΔAICc between the Use and Ignore Uncertainty models, where a positive value indicates that the Use Uncertainty model is favored. The vertical grey line indicates the median across participants, and the shaded region illustrates the 95% bootstrapped confidence interval of the median. Only ΔAICcs are illustrated because the two model comparison metrics gave similar results.

6 Model Variants

While the Use Uncertainty model provides a good fit to the data, the two models we have considered thus far contain assumptions that could be modified. In this section, we factorially compare different formulations of the encoding, inference, and decision stage of the model (van den Berg, Awh, & Ma, 2014; Acerbi, Wolpert, & Vijayakumar, 2012; Keshvari et al., 2012). A factorial model comparison is an effective way of testing which assumptions we made were critical for accounting for human behavior, and thus which are reasonable to make conclusions about. In this section, we demonstrate that our general conclusions about the use of uncertainty do not depend on the specific assumptions we made when defining our model. We only discuss the results of the Line condition here, since it is the only condition that investigates the maintenance of uncertainty in working memory. However, we did the same analysis to the Ellipse condition data and found consistent results (Appendix 8.3).

6.1 Encoding

In both the Use and Ignore Uncertainty models, we assumed that observers' encoding noise followed that of a Variable Precision model (van den Berg et al., 2012; Fougnie et al., 2012). Here, we also consider that observers' memory precision varies only based on stimulus type, and does not fluctuate on an item-to-item basis. With this "Fixed Precision" assumption of encoding noise, the κ for each item is determined only by its stimulus type; high-reliability ellipses would be encoded with parameter κ_{high}, low-reliability ellipses with κ_{low}, and lines with κ_{line}.

6.2 Inference

Observers calculate the decision variable according to some inference process, which we allow to be independent of the true generative process. The potential model mismatch (Orhan & Jacobs, 2014; Beck et al., 2012; Acerbi et al., 2014) between the true and believed generative process could be due to a result of wrong beliefs about the generative process or computation limitations that prevent accurate representation of the generative model. We consider that observers may use partial knowledge of uncertainty, rather than fully Using or Ignoring uncertainty.

We consider that the observer may have one of four inference models, listed below in decreasing order of how many factors the observer takes into account in their uncertainty:

1. Variable precision (V): the observer believes that mean memory precision varies with the exact stimulus shape (low-reliability ellipse, high-reliability ellipse, line) and that there is additional noise for each item at each presentation. This inference model is optimal when the true generative process is Variable precision.

2. Fixed precision (F): the observer believes that memory precision varies with the exact stimulus shape (low-reliability ellipse, high-reliability ellipse, line), but does not consider that there is additional noise for each item at each presentation. This inference model is suboptimal when the true generative process is Variable precision, but optimal when the true generative process is Fixed precision.

3. Limited (L): the observer believes that memory precision varies across shapes (ellipse vs. line). This observer does not consider differences in precision between high- and low-reliability ellipses or additional noise for each item at each presentation. This observer is suboptimal.

4. Same precision (S): the observer believes that memory precision is the same throughout the condition, and does not vary with stimulus shape or anything else. This is the "Ignore Uncertainty" observer and is suboptimal.

Note that the Variable and Same precision inference schemes here are identical to that of Keshvari and others' (2012), and the Fixed precision here is equivalent to their "Equal" precision inference scheme.

6.3 Decision Rule

The Use and Ignore Uncertainty models use the optimal decision rule (Eq. 5). Note that participants may have incorrect assumptions about the noise in their memory, but still be acting in accordance with Bayesian decision theory (i.e., still using the correct decision rule), resulting in "imperfectly optimal observers" (Maloney & Zhang, 2010). Alternatively, participants could be calculating the optimal decision variable, but be using a suboptimal decision rule. Here, we consider observers who use the max rule, responding Change whenever the maximum evidence of change is greater than some criterion, k,

$$\max_i d_i > k, \tag{6}$$

rather than averaging d_is. These observers are not Bayes-optimal, but are still using probabilistic computation (i.e., still using their uncertainty) in the calculation of d_i. In fact, in many cases these decision rules do not result in substantially different behavior (Ma, Shen, Dziugaite, & van den Berg, 2015). For example, if all d_is are similar, then a max and an average will result in similar values. If the maximum d_i is substantially larger than the others, both decision rules can result in similar behavior by adjusting k.

6.4 Parameters

There are two possible encoding schemes ((V)ariable, (F)ixed), four possible inference schemes ((V)ariable, (F)ixed, (L)imited, (S)ame), and two possible decision rules ((O)ptimal, (M)ax). Factorially combining each of these characteristics would yield 16 different models. We choose not to consider the models in which the generative model is "F" but the observer assumes "V" under the assumption that people tend not to assume the (perceptual) world is more complicated than it actually is; thus, we test a total of 14 models. We denote each model by the letters corresponding to their encoding scheme, inference scheme, and decision rule (e.g., VVO is the model with Variable precision encoding, an observer assumes Variable precision, and an Optimal decision rule). The VVO model is the Use Uncertainty model; the VSO model is the Ignore Uncertainty model.

Encoding parameters. Like before, observers with Variable precision encoding have parameters $\bar{J}_{\text{high}}$, $\bar{J}_{\text{low}}$, $\bar{J}_{\text{line}}$, and τ. Observers with Fixed precision encoding have parameters J_{high}, J_{low}, and J_{line}.

Inference parameters. For the observer who correctly infers their encoding process (i.e., VVO, VVM, FFO, or FFM), there are no additional parameters. If the observer has Variable precision encoding but does not take into account individual-item variations (i.e., VFO or VFM), then the assumed precision is $J_{\text{high}} = \bar{J}_{\text{high}}$, $J_{\text{low}} = \bar{J}_{\text{low}}$, and $J_{\text{line}} = \bar{J}_{\text{line}}$ for high-reliability ellipses, low-reliability ellipses, and lines, respectively. Limited inference observers (i.e., VLO, VLM, FLO, FLM) have two additional parameters: $J_{\text{ass, e}}$ and $J_{\text{ass, l}}$, corresponding to the assumed precision of the ellipses and lines, respectively. Same inference observers, who do not take any memory variations into account (i.e., VSO, VSM, FSO, FSM), have one additional parameter J_{ass}, corresponding to the assumed precision of all items.

Decision parameters. Observers using both the optimal or max decision rule have parameter k, corresponding to the decision criterion. If any item has a decision variable greater than k, then they will respond Change.

Each model and their corresponding parameters is listed in Table 2. Note that the Same inference observer who uses the max rule (i.e., VSM, FSM) has one less parameter than their Optimal decision rule counterpart (i.e., VSO, FSO) because making a decision depends only on the item with the largest measured change.

Encoding	Inference	Decision Rule	
		(O)ptimal	(M)ax
(V)ariable	(V)ariable	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, k, \lambda, \sigma_{\text{d}}(, \bar{J}_{\text{line}})$	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, k, \lambda, \sigma_{\text{d}}(, \bar{J}_{\text{line}})$
	(F)ixed	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, k, \lambda, \sigma_{\text{d}}(, \bar{J}_{\text{line}})$	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, k, \lambda, \sigma_{\text{d}}(, \bar{J}_{\text{line}})$
	(L)imited	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \bar{J}_{\text{ass,e}}\tau, k, \lambda, \sigma_{\text{d}}(, \bar{J}_{\text{line}}, \bar{J}_{\text{ass,l}})$	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \bar{J}_{\text{ass,e}}\tau, k, \lambda, \sigma_{\text{d}}(, \bar{J}_{\text{line}}, \bar{J}_{\text{ass,l}})$
	(S)ame	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \bar{J}_{\text{ass}}, \tau, k, \lambda, \sigma_{\text{d}}(, \bar{J}_{\text{line}})$	$\bar{J}_{\text{high}}, \bar{J}_{\text{low}}, \tau, k, \lambda, \sigma_{\text{d}}(, \bar{J}_{\text{line}})$
(F)ixed	(F)ixed	$J_{\text{high}}, J_{\text{low}}, k, \lambda, \sigma_{\text{d}}(, J_{\text{line}})$	$J_{\text{high}}, J_{\text{low}}, k, \lambda, \sigma_{\text{d}}(, J_{\text{line}})$
	(L)imited	$J_{\text{high}}, J_{\text{low}}, J_{\text{ass,e}}, k, \lambda, \sigma_{\text{d}}(, J_{\text{line}}, J_{\text{ass,l}})$	$J_{\text{high}}, J_{\text{low}}, J_{\text{ass,e}}, k, \lambda, \sigma_{\text{d}}(, J_{\text{line}}, J_{\text{ass,l}})$
	(S)ame	$J_{\text{high}}, J_{\text{low}}, J_{\text{ass}}, k, \lambda, \sigma_{\text{d}}(, J_{\text{line}})$	$J_{\text{high}}, J_{\text{low}}, k, \lambda, \sigma_{\text{d}}(, J_{\text{line}})$

Table 2: **Model parameters**. Model parameters for Line condition. Parameters unnecessary for fitting the Ellipse condition are displayed in parentheses. The top colored cell corresponds to parameters of the Use Uncertainty (VVO) model. The bottom colored cell corresponds to the parameters of the Ignore Uncertainty (VSO) model.

6.5 Model Comparison Results

As previously described, we estimated parameters for each participant and compared models using AICc and BIC. In this section, we only discuss the results of the Line condition using median AICc and BIC differences between the VVO (Use Uncertainty) and other models. However, we report the results of the Ellipse condition in the Appendix 8.3.

The factorial model comparison results corroborate our earlier result. Models in which observers correctly used their memory uncertainty (i.e., VVO, VVM, FFO, FFM) were indistinguishable from one another according to ΔAICc and ΔBIC (Table 3, bar graphs in Figure 5). These four models, however, fit substantially better than the remaining 10 models in which the observer does not fully know or use their memory uncertainty in their decisions, indicated by the 95% CI of the median ΔAICc and ΔBIC being greater than 0. These results imply that the specific encoding, inference, and decision rule do not matter as much the match between the encoding and inference scheme. In other words, the models that are able to capture the data are those in which observers' uncertainty reflects that of the true encoding process.

In the Appendix, we additionally report the sum of the ΔAICc and ΔBIC and group Bayesian Model Selection (BMS) for both conditions (Appendix 8.4 and 8.5, respectively). Summing the ΔAICc and ΔBIC explicitly assumes that participants are all fit by the same model, while group BMS allows for participant heterogeneity and directly infers the distribution of participants across models. Using these alternative model comparison metrics do not change the results; behavior was overwhelmingly best explained by a model that assumes that uncertainty is maintained accurately and used.

		Decision Rule			
Encoding	Inference	(O)ptimal		(M)ax	
		ΔAICc	ΔBIC	ΔAICc	ΔBIC
(V)ariable	(V)ariable	0 [0, 0]	0 [0, 0]	4 [-1, 13]	4 [-1, 13]
	(F)ixed	11 [4, 26]	11 [4, 26]	22 [11, 50]	22 [11, 50]
	(L)imited	97 [43, 256]	123 [69, 283]	255 [30, 366]	282 [56, 392]
	(S)ame	108 [58, 370]	122 [71, 383]	170 [138, 235]	170 [138, 235]
(F)ixed	(F)ixed	3 [3, 19]	-10 [-16, 9]	19 [9, 38]	6 [-4, 25]
	(L)imited	53 [23, 82]	66 [37, 96]	51 [24, 100]	64 [37, 113]
	(S)ame	39 [17, 91]	39 [17, 91]	148 [84, 167]	135 [70, 154]

Table 3: **Median ΔAICc and ΔBIC: Line condition**. The median and 95% bootstrapped confidence interval of the ΔAICc and ΔBIC. A positive value indicates that the VVO model provides a better fit to the data. The cells corresponding to the Use (VVO) and Ignore (VSO) Uncertainty models are colored in blue.

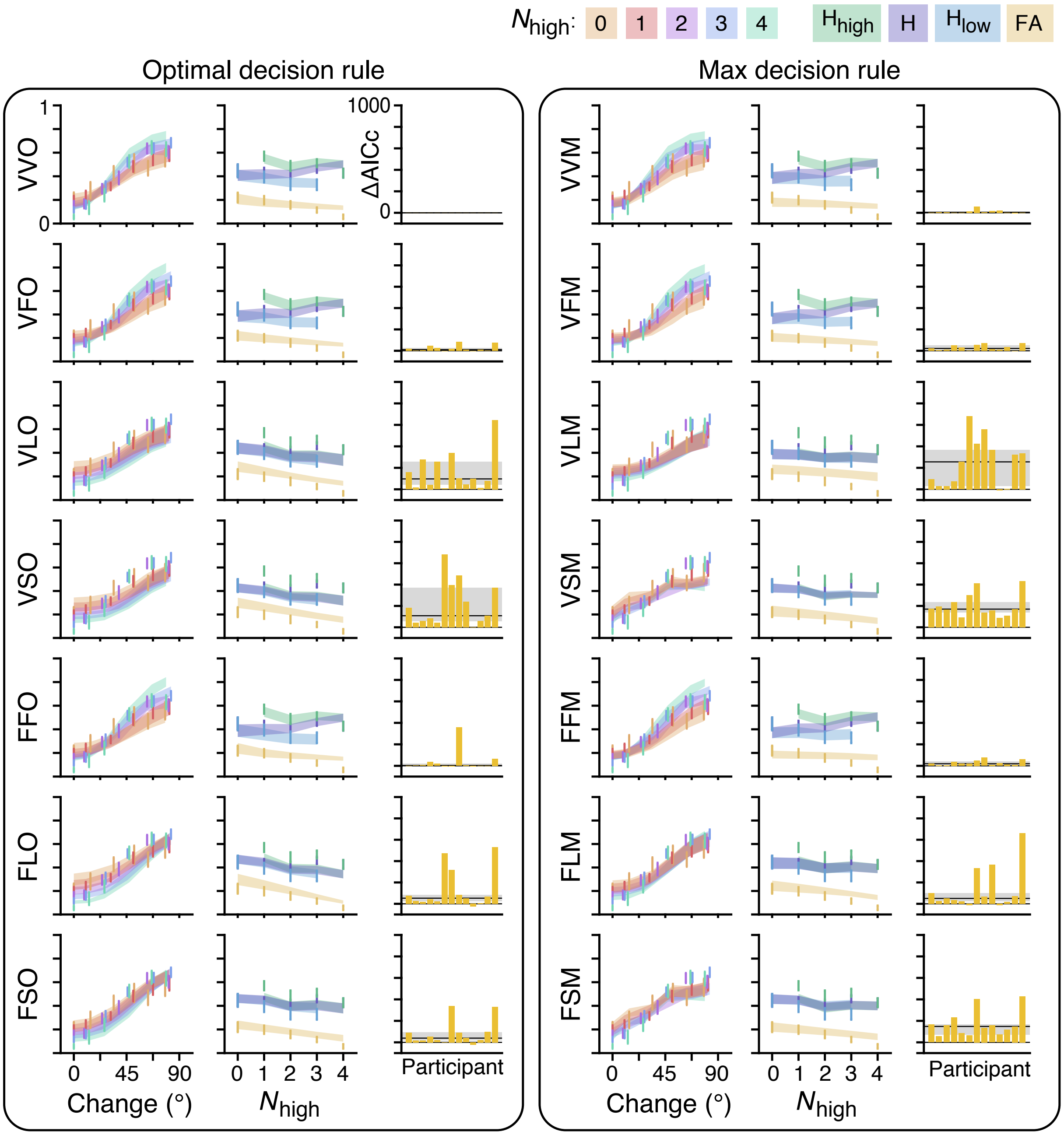

Figure 5: **Factorial model comparison.** Model predictions and performance of all possible combinations of different encoding, inference, and decision rules. $M \pm SEM$ data (error bars) and model fits (fills) for all models, organized into two columns by decision rule. For each model (each row within each column), the left graph illustrates the proportion report change as a function of amount of change. Color indicates the number of high-reliability ellipses (legend at the top of the figure). The middle graph illustrates the proportion hits for high-reliability items (green), hits for low-reliability items (blue), hits averaged across the display (purple), and false alarms (gold) as a function of number of high-reliability items (legend at the top right of the figure). The right graph illustrates the individual-participant ΔAICc, where positive numbers indicate the VVO model is a better fit to the data. The grey horizontal line and shaded region illustrates median and the 95% bootstrapped confidence interval of the median across participants.

7 Discussion

In this paper, we investigated whether uncertainty is maintained and implicitly used in a working memory-based decision. First, we demonstrated that people use uncertainty implicitly in a working memory task if that uncertainty information was available after the delay (i.e., if uncertainty did not need to be maintained). Second, and more importantly, we showed that people not only use uncertainty, but maintain this information over the working memory delay. Finally, we factorially tested different model encoding schemes, inference schemes, and decision rules and found that people were best described by models in which observers accurately maintain and use uncertainty in their decision.

First, we demonstrated that people could use uncertainty implicitly in a working memory task if that uncertainty information was experimentally available. While the change detection task has been an experimental staple in the working memory literature (e.g., Luck & Vogel, 1997; Phillips, 1974; Pashler, 1988), the majority of these tasks feature large, categorical changes in the stimulus. In contrast, our task, which is a direct experimental replication of that of Keshvari and others (2012), featured changes that varied on a trial-to-trial basis. Trial-to-trial fluctuations in stimuli and witholding of feedback allow for a strongest test of probabilistic computation because observers would need to maintain a belief distribution over stimulus values to maximize performance in this task (Ma & Jazayeri, 2014). Through formal model comparison, we showed that all participants in the Ellipse condition are better fit by the Use Uncertainty model than the Ignore Uncertainty model. The Use Uncertainty model was identical to the model that was found to describe participant data best in the study by Keshvari et al. (2012). These results are also qualitatively consistent with Devkar and others' (2017) work, despite being slightly different tasks.

Second, and more importantly, we showed that people not only use uncertainty, but maintain this information over the working memory delay. Like in the Ellipse condition, we found that all participants in the Line condition were better fit by the Use Uncertainty model than the Ignore Uncertainty model. However, the conclusion of this model comparison is critically different. In the Ellipse condition as well as in previous studies (Keshvari et al., 2012; Devkar et al., 2017), the ellipses were presented after the working memory delay, with the same reliability as before. With these experimental designs, reliability information could be used as a heuristic to inform uncertainty, thus not requiring this information to be maintained in memory. In other words, these previous studies cannot make any conclusions about the contents of working memory, only the decision-making process that follows it. Our result, in contrast, demonstrates that uncertainty was actually *maintained* in working memory, since the information was not available to the participants through a heuristic such as ellipse reliability.

These results provide evidence that people can maintain and use uncertainty in an *implicit* decision-making paradigm. This is in agreement with studies that asked participants to make an explicit reports such as confidence ratings (Rademaker et al., 2012; Vandenbroucke et al., 2014; Samaha & Postle, 2017). This is also in agreement with the "choose best" (Fougnie et al., 2012; Suchow et al., 2017) and wager paradigms (Yoo et al., 2018; Honig et al., 2020). These paradigms are explicit in the sense that partipants deliberate on a decision that is directly related to their conscious uncertainty in a memory, but implicit in the sense that this uncertainty must be taken into account in another perceptual decision.

Finally, we conducted a factorial model comparison to investigate whether our conclusions were due to specific assumptions about model encoding schemes, inference schemes, and decision rules. We found that the most important aspect of a model was that its encoding and inference schemes were matched, providing strong evidence that people accurately maintain and use their uncertainty in working memory-based decisions. Our results are somewhat inconsistent with research showing strong support for Variable precision encoding over Fixed precision encoding (van den Berg et al., 2014; Keshvari et al., 2012), and consistent with research showing that optimal and max decision rules are difficult to distinguish through model comparison (Ma et al., 2015). Future experiments may attempt to tease apart these two decision rules by manipulating the prior probability of change trials or the behavioral relevance of individual items.

Our results suggest that existing computational models of working memory that currently ignore uncertainty should be updated. For example, attractor network models currently maintain a point estimate of a single item feature through the mean of a stereotyped bump in a network of neurons (Ermentrout, 1998; Wang, 2001; Compte, 2006). Thus, there is typically no notion of uncertainty in this framework. Lim and Goldman (2014) demonstrated that altering the network connectivity and dynamics results in "negative-derivative feedback models," in which networks can vary not only in mean but also in amplitude. Probabilistic population coding (PPC) and neural network models have implemented precision through input gain (Ma, Beck, Latham, & Pouget, 2006; Orhan & Ma, 2017). Additional research must investigate whether these negative-derivative feedback models can represent a memory's precision through the amplitude of the network maintaining it, precision which could be read out from the observer as uncertainty.

Additionally, computational models could be used to decode uncertainty from neural activity in working memory tasks. Work in visual perception demonstrates that uncertainty information is represented in primary visual cortex (van Bergen, Ma, Pratte, & Jehee, 2015; van Bergen, 2019; Walker, Cotton, Ma, & Tolias, 2020; Hénaff, Boundy-Singer, Meding, Ziemba, & Goris, 2020). These studies built normative Bayesian models to infer stimulus value from BOLD signal. The likelihood of the stimulus, and thus uncertainty, could be read out from the models. Estimates of trial-specific uncertainty are positively correlated with error, suggesting that primary visual cortex held uncertainty information. Since working memories have been shown to be maintained in the same sensory areas with which they are perceived (e.g. Curtis & D'Esposito, 2003; Postle, 2006; D'Esposito & Postle, 2015; Harrison & Tong, 2009), perhaps visual working memory uncertainty is also stored in visual cortex. To more rigorously test the representation of uncertainty decoded from BOLD data, future studies can correlate decoded uncertainty with behavioral measures of uncertainty such as confidence ratings (Rademaker et al., 2012) or post-decision wagers (Yoo et al., 2018; Honig et al., 2020). Additionally, future studies can try to fit individual-trial data using these methods, which is more compelling evidence in favor of a model than a correlation.

Overall, this paper shows that people have uncertainty that reflects their memory noise at an item-specific level and they maintain this information over a working memory delay. This research demonstrates that there is other information, beyond a point estimate, maintained in working memory and used in later decisions.

Acknowledgements: We thank Marissa Evans for collecting the bulk of the data in this study and Emin Orhan for collaborating on a previous iteration of this project. W.J.M is supported by award number R01EY020958. A.H.Y. was supported by training grant T32 EY7136-25. This work was supported in part through the NYU IT High Performance Computing resources, services, and staff expertise.

References

Acerbi, L., & Ma, W. J. (2017). Practical Bayesian optimization for model fitting with Bayesian Adaptive Direct Search. *Advances in Neural Information Processing Systems*, *30*, 1834–1844.

Acerbi, L., Ma, W. J., & Vijayakumar, S. (2014). A framework for testing identifiability of Bayesian models of perception. *Advances in Neural Information Processing Systems*, 1026–1034.

Acerbi, L., Wolpert, D. M., & Vijayakumar, S. (2012). Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLOS Computational Biology*, *8*(11).

Adam, K. C. S., & Vogel, E. K. (2017). Confident failures: Lapses of working memory reveal a metacognitive blind spot. *Attention, Perception & Psychophysics*, *79*(5), 1506–1523. doi: 10.3758/s13414-017-1331-8

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262. doi: 10.1016/j.cub.2004.01.029

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839. doi: 10.1038/nrn1201

Baddeley, A., & Hitch, G. (1974). Working memory. *The Psychology of Learning and Motivation*, *8*, 47–89.

Barthelme, S., & Mamassian, P. (2010). Flexible mechanisms underlie the evaluation of visual confidence. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(48), 20834–20839. doi: 10.1073/pnas.1007704107

Bays, M. P., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854.

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not noisy, just wrong: The role of suboptimal inference in behavioral variability. *Neuron*, *74*(1), 30–39. doi: 10.1016/j.neuron.2012.03.016

Bona, S., Cattaneo, Z., Vecchi, T., Soto, D., & Silvanto, J. (2013). Metacognition of visual short-term memory: Dissociation between objective and subjective components of vstm. *Frontiers in Psychology*, *4*(62).

Bona, S., & Silvanto, J. (2014). Accuracy and confidence of visual short-term memory do not go hand-in-hand: Behavioral and neural dissociations. *PLOS One*, *9*(3). doi: 10.1371/journal.pone.0090808

Compte, A. (2006). Computational and in vitro studies of persistent activity: Edging towards cellular and synaptic mechanisms of working memory. *Neuroscience*, *139*(1), 135–151.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. doi: 10.1016/j.tics.2003.10.005

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, *7*(9), 415–423. doi: 10.1016/S1364-6613(03)00197-9

D'Esposito, M., & Postle, R. B. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142. doi: 10.1146/annurev-psych-010814-015031

Devkar, D., Wright, A. A., & Ma, W. J. (2017). Monkeys and humans take local uncertainty into account when localizing a change. *Journal of Vision*, *17*(11). doi: 10.1167/17.11.4

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380–417.

Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Reports on Progress in Physics*, *61*(4), 353–430. doi: 10.1088/0034-4885/61/4/002

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. doi: 10.1038/415429a

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*(5), 673–679. doi: 10.3758/17.5.673

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. doi: 10.1038/nature07832

Hénaff, O. J., Boundy-Singer, Z. M., Meding, K., Ziemba, C. M., & Goris, R. L. T. (2020). Representation of visual uncertainty through neural gain variability. *Nature Communications*, *11*.

Honig, M., Ma, W. J., & Fougnie, D. (2020). Humans incorporate trial-to-trial working memory uncertainty into rewarded decisions. *Proceedings of the National Academy of Sciences*, *117*(15), 8391–8397. doi: 10.1073/pnas.1918143117

Hurvich, C. M., & Tsai, C. L. (1987). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307.

Jazayeri, M., & Shadlen, N. M. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149. doi: 10.1037/0033-295X.99.1.122

Keshvari, S., van den Berg, R., & Ma, W. J. (2012). Probabilistic computation in human perception under variability in encoding precision. *PLOS One*, *7*.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. doi: 10.1016/j.tins.2004.10.007

Körding, P. K., & Wolpert, M. D. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244–7.

Levenstein, D., Alvarez, V. A., Amarasingham, A., Azab, H., Gerkin, R. C., Hasenstaub, A., … Redish, D. A. (2020). On the role of theory and modeling in neuroscience. *arXiv*.

Lim, S., & Goldman, M. S. (2014). Balanced cortical microcircuitry for spatial working memory based on corrective feedback control. *Journal of Neuroscience*, *34*(20), 6790–6806. doi: 10.1523/JNEUROSCI.4602-13.2014

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Ma, W. J. (2010). Signal detection theory, uncertainty, and Poisson-like population codes. *Vision Research*, *50*(22), 2308–2319. doi: 10.1016/j.visres.2010.08.035

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432–1438. doi: 10.1038/nn1790

Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual review of neuroscience*, *37*, 205–220. doi: 10.1146/annurev-neuro-071013-014017

Ma, W. J., Navalpakkam, V., Beck, M. J., van den Berg, R., & Pouget, A. (2011). Behavior and neural basis of near-optimal visual search. *Nature Neuroscience*, *14*(6), 783–90.

Ma, W. J., Shen, S., Dziugaite, G., & van den Berg, R. (2015). Requiem for the max rule? *Vision Research*, *116*, 179–193.

Maloney, L. T., & Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual Neuroscience*, *26*(1), 147–155. doi: 10.1017/S0952523808080905

Maloney, L. T., & Zhang, H. (2010). Decision-theoretic models of visual perception and action. *Vision Research*, *50*(23), 2362–2374. doi: 10.1016/j.visres.2010.09.031

Maniscalco, B., & Lau, H. (2015, Jul 3). Manipulation of working memory contents selectively impairs metacognitive sensitivity in a concurrent visual discrimination task. *Neuroscience of consciousness*, *2015*(1). doi: 10.1093/nc/niv002

Mueller, S. T., & Weidemann, C. T. (2008). Decision noise:An explanation for observed violations of signal detection theory. *Psychonomic Bulletin & Review*, *15*(3), 465–494. doi: 10.3758/PBR.15.3.465

Orhan, A. E., & Jacobs, R. A. (2014). Are performance limitations in visual short-term memory tasks due to capacity limitations or model mismatch? *arXiv*.

Orhan, A. E., & Ma, W. J. (2017). Efficient probabilistic inference in generic neural networks traiined with non-probabalistic feedback. *Nature Communications*, *8*(138).

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378.

Phillips, A. W. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2). doi: 10.3758/BF03203943

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23–38. doi: 10.1016/j.neuroscience.2005.06.005

Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision*, *12*, 21–21.

Sahar, T., Sidi, Y., & Makovski, T. (2020). A metacognitive perspective of visual working memory with rich complex objects. *Frontiers in Psychology*, *11*.

Samaha, J., Barrett, J. J., Sheldon, D. A., LaRocque, J. J., & Postle, R. B. (2016). Dissociating perceptual confidence from discrimination accuracy reveals no influence of metacognitive awareness on working memory. *Frontiers in Psychology*, *7*.

Samaha, J., & Postle, B. R. (2017). Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. *Proceedings of The Royal Society B Biological Sciences*, *284*(1867). doi: 10.1098/rspb.2017.2035

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

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*, 1004–1017. doi: 10.1016/j.neuroimage.2009.03.025

Stocker, A. A., & Simoncelli, P. E. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neurocience*, *9*(4), 578–85.

Suchow, J. W., Fougnie, D., & Alvarez, G. A. (2017). Looking inward and back: Real-time monitoring of visual working memories. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(4), 660–668. doi: 10.1037/xlm0000320

van Beers, J. R., Sittig, C. A., & Gon, J. J. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of neurophysiology*, *81*(3), 1355–64.

van Bergen, R. S. (2019). Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. *Journal of Neuroscience*, *39*(41), 8164–8176. doi: 10.1523/JNEUROSCI.3212-18.2019

van Bergen, R. S., Ma, W. J., Pratte, M. S., & Jehee, J. F. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience*(18), 1728–1730.

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*, 124–149. doi: 10.1037/a0035234

van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*, 8780–8785.

van den Berg, R., Yoo, A. H., & Ma, W. J. (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. *Psychological Review*, *124*(2), 197–214. doi: 10.1037/rev0000060

Vandenbroucke, E. A. R., Sligte, I. G., Barrett, A. B., Seth, A. K., Fahrenfort, J. J., & Lamme, V. A. (2014). Accurate metacognition for visual sensory memory representations. *Psychological science*, *25*(4).

van Opheusden, B., Acerbi, L., & Ma, W. J. (2020). Unbiased and efficient log-likelihood estimation with Inverse Binomial Sampling. *arXiv*.

Vlassova, A., Donkin, C., & Pearson, J. (2014, Nov 11). Unconscious information changes decision accuracy but not confidence. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(45), 16214–16218. doi: 10.1073/pnas.1403619111

Walker, E., Cotton, J., Ma, W. J., & Tolias, A. (2020). A neural basis of probabilistic computation in visual cortex. *Nature Neuroscience*, *23*, 122–129. doi: 10.1038/s41593-019-0554-5

Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosciences*, *24*(8), 455–463.

Yoo, A. H., Klyszejko, Z., Curtis, E. C., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports*, *8*(1). doi: 10.1038/s41598-018-34282-1

Zhang, W., & Luck, J. S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.