The two-alternative forced-choice (2AFC) task is the workhorse of psychophysics and is used to measure the just-noticeable difference, generally assumed to accurately quantify sensory precision. However, this assumption is not true for all mechanisms of decision making. Here we derive the behavioral predictions for two popular mechanisms, sampling and maximum a posteriori, and examine how they affect the outcome of the 2AFC task. These predictions are used in a combined visual 2AFC and estimation experiment. Our results strongly suggest that subjects use a maximum a posteriori mechanism. Further, our derivations and experimental paradigm establish the already standard 2AFC task as a behavioral tool for measuring how humans make decisions under uncertainty.

Introduction

Virtually all our decisions are made in the presence of uncertainty. For example, say we are choosing which checkout line at the grocery store to wait in. We have our past experience to inform us which lines usually move fast (prior beliefs). We also have our observations of the current situation (e.g., the number of people in each line), which give us an estimate of how fast each line will move (the likelihood). By combining these two pieces of information, a belief about which line is fastest (a posterior belief) can be formed. Many studies have shown, under a broad set of conditions, that people do combine a prior belief and a likelihood to make decisions under uncertainty. Though decisions such as these are commonplace, exactly how the brain
uses these posterior beliefs to choose between multiple uncertain options remains largely unknown.

Several theories propose mechanisms by which the brain uses uncertain beliefs to make a choice. In the checkout line example, our beliefs for each line are represented as a probability distribution over wait times. One theory proposes that the brain chooses the line for which the most probable wait time is the shortest. This is called the maximizing or maximum a posteriori (MAP) choice (Green, 1966). An alternative theory proposes that the brain cannot compute the MAP but can instead get samples from these believed waiting-time distributions. By using many such samples, the brain can estimate the expected wait times for each line and make a choice. This is called the sampling hypothesis (Ackley, Hinton, & Sejnowski, 1985; Hoyer & Hyvärinen, 2003). Both MAP and sampling can be viewed as normative models of decision making (Sakai & Fukai, 2008b; Vul, Goodman, Griffiths, & Tenenbaum, 2009; Wozny et al., 2010; Duda, Hart, & Stork, 2012), making either an informative candidate for how the brain makes decisions.

To examine decision making in a broad range of disciplines, studies have used one of the simplest decision tasks: the two-alternative forced-choice (2AFC) task (Green, 1966). In this task, subjects are asked to choose one of two possible options. For example, subjects may be asked which of two checkout lines is fastest, which of two tones has a higher pitch, or which of two flashes of light displayed on a screen is farther to the left. By controlling the discrepancy between these stimuli, experimenters can obtain a psychometric curve: the probability of a subject’s response given the discrepancy between stimuli. This curve can then be used to measure the just-noticeable difference (JND), which quantifies how different the two stimuli must be before subjects reliably report them as distinct. As we demonstrate later in this article, even though the 2AFC task has been the workhorse of psychophysics for the past 150 years (Green, 1966), interpreting the JND relies on the still-unknown mechanisms of decision making.

Interestingly, the measured JND is almost always interpreted as characterizing sensory precision (e.g., Ernst & Banks, 2002; Stocker & Simoncelli, 2006; Tassinari, Hudson, & Landy, 2006; Fetsch, Pouget, DeAngelis, & Angelaki, 2011; Girshick, Landy, & Simoncelli, 2011) and is thought to be independent of an individual’s prior beliefs. Depending on the mechanism of decision making, however, the JND may or may not be influenced by a subject’s prior, thus confounding the JND’s interpretation. This confound raises concerns for the great number of studies that rely on the 2AFC task to measure sensory precision. Yet it also suggests that the JND’s dependence on the prior could be used as a tool for probing the decision-making mechanism of the brain.

Here we derive the theoretical predictions for behavior in the 2AFC task for multiple mechanisms, including MAP and sampling. We demonstrate how only in the maximizing case is there no influence of the prior on the JND, in which case it correctly quantifies sensory uncertainty. We then exploit this result to design an experimental paradigm that measures JNDS as a function of prior uncertainty. Using an interleaved estimation and 2AFC task, we measure subjects’ prior beliefs and their JNDS. We found that changes in a subject’s prior had no measurable influence on their JND, consistent with MAP decision making. Additionally, we establish a new use of the 2AFC task to probe the mechanisms of decision making.

Materials and methods

Theoretical prediction of psychometric curves

The 2AFC behavior is a two-stage decision process: forming a belief about the stimuli’s values and making a decision based on these beliefs (Figure 1A). Formally, the sensory information about the stimuli (i.e., the cues) induces a distribution over their possible values (i.e., a likelihood). Subjects form a belief (posterior distribution) in stimulus values by combining their prior expectations with the likelihoods. This Bayesian belief-formation process is well supported by many studies (Yuille & Bülthoff, 1994; Knill & Richards, 1996; Ernst & Banks, 2002; Kording & Wolpert, 2004; Berniker, Voss, & Kording, 2010; Vilares, Howard, Fernandes, Gottfried, & Kording, 2012). How subjects make a decision based on these beliefs is less well established. Here we begin by formally defining the 2AFC task in terms of subjects’ beliefs, and then we mathematically examine the results under the assumption of two popular classes of decision making. We highlight the differences between these decision-making mechanisms and how experiments might test for them.

Decision making and the 2AFC task

In the 2AFC task (Green, 1966), subjects make a binary choice based on two experimentally imposed stimuli, which we refer to as \( s_1 \) and \( s_2 \). We denote the subject’s response as \( z \) (\( z = 0 \) if false, \( z = 1 \) if true). By asking the subject to perform many trials while systematically manipulating the disparity \( \delta = s_2 - s_1 \), we can measure a psychometric curve, characterizing the probability of a subject’s decision as a function of the disparity. As is standard, we fit this curve with the
cumulative normal distribution:

\[ P(z = 1|\delta) = \frac{1}{2} \left[ 1 + \text{erf}\left(\frac{\delta - \text{PSE}}{\sqrt{2}\sigma_{\text{JND}}}\right) \right] \]

where \( \text{erf} \) denotes the error function and PSE is the probability of subjective equality: the subjective bias in stimulus disparity. One of the central measures obtained from fitting the psychometric curve is the JND, which quantifies how different the stimuli must be for subjects to reliably discriminate between them. Though the definition of JND varies slightly across studies, here we define it to be the best fit of \( \sigma_{\text{JND}} \) to behavior. This JND is of central importance across much of psychophysics.

Due to experimental, sensory, and neural sources of noise, subjects cannot observe the true stimuli values \( s_1 \) and \( s_2 \) and instead have to rely on noisy sensations of the cues, \( c_1 \) and \( c_2 \). Formally, our sensory information induces the likelihood of all possible values of the stimulus given the cues, \( P(c_1|s_1) \) and \( P(c_2|s_2) \). Subjects’ percepts can be thought of as a belief distribution over stimulus values given the cue, which we can describe as a posterior distribution, \( P(s|c) \). Applying Bayes’ formula, we have

\[ P(s|c) \propto P(c|s)P(s) \]

where \( P(s) \) is the subject’s prior expectation of stimulus values. Now we can formally interpret the 2AFC task as a decision, \( z \), based on beliefs about the stimulus values, \( P(s_1|c_1) \) and \( P(s_2|c_2) \). This allows us to predict different distributions, \( P(z|\delta) \), for different candidate decision-making mechanisms.

MAP decision making

According to the MAP hypothesis, subjects choose \( z \) to maximize the probability of being correct. Mathematically, the choice is defined as follows:

\[ z = \begin{cases} 
1 & \text{if } \text{argmax}_{s_2}[P(s_2|c_2)] - \text{argmax}_{s_1}[P(s_1|c_1)] \geq 0 \\
0 & \text{otherwise}
\end{cases} \]
Let us assume Gaussian distributions for the likelihoods \( P(c_1|s_1) = N(s_1, \sigma^2_1) \) and \( P(c_2|s_2) = N(s_2, \sigma^2_2) \) and the prior \( P(s) = N(\mu, \sigma^2) \). Then we can rewrite the choice as follows:

\[
z = \begin{cases} 
1 & \text{if } \delta_{MAP} \geq 0 \\
0 & \text{otherwise} 
\end{cases}
\]

(4)

where \( \delta_{MAP} \) is the difference between the expected posterior beliefs in the two stimuli:

\[
\delta_{MAP} = E[P(s_2|c_2)] - E[P(s_1|c_1)]
\]

\[
= \frac{\mu \sigma^2_2 + c_2 \sigma^2_2}{\sigma^2_2 + \sigma^2_s} - \frac{\mu \sigma^2_1 + c_1 \sigma^2_1}{\sigma^2_1 + \sigma^2_s}
\]

Therefore, a subject’s response is determined through the random variable, \( \delta_{MAP} \), which is defined by the two random cues, \( c_1 \), and \( c_2 \). Next, using \( P(c_1|s_1) \) and \( P(c_2|s_2) \), we can integrate out \( c_1 \) and \( c_2 \) to obtain the probability distribution for \( \delta_{MAP} \) in terms of the two experimental variables, \( s_1 \), \( s_2 \):

\[
P(\delta_{MAP}|s_1, s_2) = N \left( s_2 - s_1 + \frac{(s_1 - \mu) \sigma^2_1}{\sigma^2_1 + \sigma^2_s} + \frac{(\mu - s_2) \sigma^2_2}{\sigma^2_2 + \sigma^2_s}, \right.
\]

\[
\left. \sigma^2_s \sqrt{\frac{\sigma^2_1}{(\sigma^2_1 + \sigma^2_s)^2} + \frac{\sigma^2_2}{(\sigma^2_2 + \sigma^2_s)^2}} \right)
\]

(5)

Finally, we compute the probability of a subject’s response:

\[
P(z = 1|s_1, s_2) = P(\delta_{MAP} \geq 0|s_1, s_2)
\]

\[
= \int_{-\infty}^{\infty} P(\delta_{MAP}|s_1, s_2) d\delta_{MAP}
\]

\[
= \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\mu(\sigma^2_2 - \sigma^2_1) + s_2(\sigma^2_1 + \sigma^2_s) - s_1(\sigma^2_2 + \sigma^2_s)}{\sqrt{2} \sigma \sqrt{\sigma^2_1(\sigma^2_2 + \sigma^2_s)^2 + \sigma^2_2(\sigma^2_1 + \sigma^2_s)^2}} \right) \right]
\]

(6)

Equation 6 defines the psychometric curve, or the probability of the subject response given the experimentally manipulated stimuli. First we note a few important features of this curve. In this general case, where the two sensed stimuli have different likelihoods, the psychometric curve is a function of the two stimuli, \( s_1 \) and \( s_2 \), and cannot be rewritten in terms of their difference, \( s_2 - s_1 \); that is, the psychometric curve is actually a surface. Also note that the point of subjective equality (PSE) is found when

\[
P(z = 1|s_1, s_2) = \frac{1}{2} \iff s_2(\sigma^2_1 + \sigma^2_s) + \mu \sigma^2_2 = s_1(\sigma^2_2 + \sigma^2_s) + \mu \sigma^2_1
\]

(7)

Therefore, the PSE varies linearly with the stimulus positions, a fact we exploit to verify that subjects are actually using their prior during experiments (explained later). The JND is found to be

\[
\sigma_{JND} = \sqrt{\sigma^2_1(\sigma^2_2 + \sigma^2_s)^2 + \sigma^2_2(\sigma^2_1 + \sigma^2_s)^2} / (\sigma^2_1 + \sigma^2_s)
\]

(8)

If, however, the two likelihoods have the same variance, \( \sigma_1 = \sigma_2 = \sigma \), then the psychometric curve collapses to the more familiar form defined by the difference of the two stimuli, \( \delta \):

\[
P(z = 1|\delta) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\delta}{2\sigma} \right) \right]
\]

(9)

To summarize, if subjects choose according to the MAP hypothesis, then Equation 9 accurately models their behavior, whereas Equation 1 is used to fit their behavior. By comparing terms with Equation 1, we can define the experimentally derived JND in terms of the precision of a subject’s likelihood:

\[
\sigma_{JND}^{MAP} = \sqrt{2}\sigma
\]

(10)

This result is implicitly used by most studies that employ the 2AFC task to measure sensory uncertainty. If these assumptions are true, then the prior indeed has no influence on 2AFC behavior or the JND, making the approach particularly attractive (Figure 1B). However, as we demonstrate later, alternative decision-making mechanisms predict distinct results.

**Sample-based decision making**

According to the sampling hypothesis, subjects choose based on approximations to the most probable stimulus values. This approximation is computed by sampling from the two posterior distributions and comparing the averages of these samples. Mathematically, we can express the choice as follows:

\[
z = \begin{cases} 
1 & \text{if } s^k_2 - s^k_1 \geq 0 \\
0 & \text{otherwise} 
\end{cases}
\]

(11)

where \( s^k_1 \) and \( s^k_2 \) are the sample means computed by drawing \( k \) samples from the posterior distributions \( P(s_1|c_1) \) and \( P(s_2|c_2) \), respectively; that is, \( s^k_i = \frac{1}{k} \sum_{i=1}^{k} s_{1i} \), where \( s_{1i} \sim P(s_1|c_1) \). Here again, if we assume Gaussian distributions for the likelihoods and prior we can rewrite the choice as follows:

\[
z = \begin{cases} 
1 & \text{if } \delta_{sample} \geq 0 \\
0 & \text{otherwise} 
\end{cases}
\]

(12)

where \( \delta_{sample} = s^k_2 - s^k_1 \), is the random variable defined by the two random cues, \( c_1 \), and \( c_2 \);
Table 1. JND predictions of decision-making theories.

\[
P(\delta_{\text{sample}}|s_1, s_2) = \mathcal{N}(\mu_{\delta}, \sigma_{\delta}^2)
\]

where the mean and variance are

\[
\mu_{\delta} = s_2 - s_1 + \frac{\sigma^2_{s_1} (\mu - s_2)}{\sigma_{s_1}^2 + \sigma_{s_2}^2} - \frac{\sigma^2_{s_2} (\mu - s_1)}{\sigma_{s_1}^2 + \sigma_{s_2}^2}
\]

and

\[
\sigma_{\delta}^2 = \frac{\sigma_{s_1}^4}{\sigma_{s_1}^2 + \sigma_{s_2}^2} + \frac{\sigma_{s_2}^4}{\sigma_{s_1}^2 + \sigma_{s_2}^2} - \frac{2 \sigma_{s_1}^2 \sigma_{s_2}^2}{\sigma_{s_1}^2 + \sigma_{s_2}^2} + \frac{\sigma_{s_2}^4}{\sigma_{s_1}^2 + \sigma_{s_2}^2}
\]

Finally, we compute the probability of a subject’s response:

\[
P(z = 1|s_1, s_2) = \int_0^\infty P(\delta_{\text{sample}}|s_1, s_2) d\delta_{\text{sample}}
\]

\[
= \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\mu_{\delta}}{\sqrt{2\sigma_{\delta}}} \right) \right]
\]

As with the MAP decision mechanism, the psychometric curve cannot be written in terms of the difference between \(s_1\) and \(s_2\). The PSE is the same as in the MAP case (Equation 7). The JND is different, however, and given by

\[
\sigma_{\text{JND}} = \frac{(\sigma_{s_1}^2 + \sigma_{s_2}^2)}{\sigma_{s_2}^2}
\]

\[
\times \sqrt{\frac{\sigma_{s_1}^4}{(\sigma_{s_1}^2 + \sigma_{s_2}^2)^2} + \frac{\sigma_{s_2}^4}{(\sigma_{s_1}^2 + \sigma_{s_2}^2)^2} + \frac{2 \sigma_{s_1}^2 \sigma_{s_2}^2}{(\sigma_{s_1}^2 + \sigma_{s_2}^2)^2} + \frac{\sigma_{s_2}^4}{(\sigma_{s_1}^2 + \sigma_{s_2}^2)^2}}
\]

\[
\times \frac{1}{2 \sigma_{s_1}^2} \left( \frac{\sqrt{k} \sigma_{s_1} \delta}{\sqrt{k \sigma_{s_2}^2 + \sigma_{s_2}^2}} \right)
\]

\[
\sigma_{\text{JND}} = \sqrt{\frac{2}{k}} \sigma_{s_2} \sqrt{\frac{k + 1}{2}}\frac{\sigma_{s_2}^2}{\sigma_{s_1}^2} + \frac{\sigma_{s_2}^2}{\sigma_{s_1}^2}
\]

Only when the two likelihoods have the same variance, \(\sigma\), does the psychometric curve collapse to a form defined by the difference of the two stimuli, \(\delta\):

\[
P(z = 1|\delta) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\sqrt{k} \sigma_{s_1} \delta}{2 \sigma_{s_2} \sqrt{(k + 1) \sigma_{s_2}^2 + \sigma_{s_2}^2}} \right) \right]
\]
This is sensible given that in our experiments subjects necessarily need to estimate the stimuli locations, for either estimating their positions or choosing which was farther to the right (see Experimental protocol later). However, it could be the case that subjects sample from their prior to compute the most likely estimate. Or, alternatively, subjects could sample from the posterior over the decision variable, \( z \). We examined this latter hypothesis (see Supplemental Appendix), and while there is no closed-form expression for the resulting distribution analogous to Equation 19, we found that this model’s influence on the JND was nearly identical to that of our model. Additionally, subjects could have a prior belief that is different from the correct Gaussian distribution in our experiment (again, see Experimental protocol). To this end we examined the influence of a uniform prior over stimuli locations. Here too we found nearly indistinguishable predictions from our Gaussian model (see Supplemental Appendix). Thus, we suggest that the results we present here for sampling may be typical for alternative interpretations.

**The PSE’s dependence on prior beliefs**

As we have shown, when the likelihoods of the two stimuli have different standard deviations, the psychometric curve and the PSE are a function of both stimulus positions, not merely their difference. For both MAP and sampling mechanisms, the PSE changes linearly with the stimulus positions. The discrepancy \( s_2 - s_1 \) that results in \( P(z = 1|s_1, s_2) = 0.5 \) is given by Equation 7. From this equation, we can obtain a relationship between either stimulus value and the PSE. For example, we can isolate the value of \( s_2 \) that corresponds to the PSE as

\[
s_2^{\text{PSE}} = \frac{s_1 + \sigma_1^2 (s_1 - \mu) + \mu \sigma_1^2}{\sigma_2^2 + \sigma_1^2}
\]

and then substitute this value to rewrite the PSE in terms of \( s_1 \) as follows:

\[
PSE(s_1) = s_2^{\text{PSE}} - s_1 = \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \mu + \frac{\sigma_2^2 - \sigma_1^2}{\sigma_1^2 + \sigma_2^2} s_1
\]  

(22)

We see how this linear relationship changes with the variance of the prior. We can exploit this relationship as a valuable control to verify that subjects use their subjective prior during our 2AFC experiments.

**Experimental protocol**

Based on the findings presented earlier in this article, we designed an experiment to examine whether 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 with a sampling mechanism. If, on the other hand, the JND is invariant with respect to their prior, this would be evidence that subjects use a MAP mechanism. Perhaps of more practical importance, this would also provide evidence that the 2AFC task measures sensory precision. To test this, we had subjects participate in a previously published “coin-catching” paradigm (Tassinari et al., 2006; Berniker et al., 2010; Vilares et al., 2012) that combined estimation and 2AFC tasks.

A virtual coin-catching paradigm was used to test subjects in both estimation and 2AFC tasks. All trials and tasks began the same way. 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 2). After 25 ms a mask was displayed for 500 ms. 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 ms and followed by a mask. After this, subjects were randomly either asked to estimate the second coin’s location (the estimation task) or asked which of the two coins landed farther to the right (the 2AFC task; see Figure 2).

In the estimation trials, subjects were presented with a virtual net that was depicted with a vertical bar (2% 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 the subjects placed the net in the desired location and pressed the mouse key, the true location of the coin was displayed and the trial ended. If a coin landed in the net, it was considered caught. A running tally of the number of coins caught and the average distance between the net and the coins were displayed. Thus, the estimation trials changed the subjects’ prior belief in coin locations.

In the 2AFC trials, subjects were instructed to guess which tossed coin, unseen to them, landed farther to the right by pressing a key (1 for the first coin, 2 for the second coin). The data collected during the 2AFC trials were used to construct psychometric curves. The curves allowed us to measure the subjects’ JNDs and to verify that the subjects used a consistent prior across the estimation and 2AFC trials.

**Experimental details**

Subjects performed the experiment over 5 days, participating approximately 2 hr per day. On each day
they were seated approximately 24 in. in front of a computer monitor (52 cm wide, 32.5 cm high) in a quiet room. Each day subjects performed two 1,000-trial blocks (for a total of 10,000 trials across 5 days). The prior over coin locations switched from block to block (e.g., from wide to narrow on one day, from narrow to wide on the subsequent day, and so on). Both priors were normally distributed with their means at the center of the screen. The narrow prior had a standard deviation of 4% screen width, whereas the wide prior had a standard deviation of 20% of screen width. To create the splashes, we sampled four dots from a Gaussian distribution centered on the hidden coin’s location using distributions with two different standard deviations (2.5% of the screen width and 10% of the screen width). These dots were resampled whenever their actual standard deviation (standard deviation of the likelihood) differed from the defined value by more than 10%. Occasionally, coins close to the left and right sides 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. Given the generative process of coins and splashes, this was expected to happen less than 1% of the time. In half of the trials, the same likelihood (8%) was used for both coins. In the remaining trials, the splashes of one of the coins used the 2.24% standard deviation, whereas the splashes of the other coin used the 8% standard deviation.

The 2AFC trials wherein the two coins had the same likelihood standard deviation were used to measure subjects’ JND. The 2AFC trials wherein the two coins had different likelihood standard deviations allowed us to verify that subjects used the prior—not simply the splash’s centroid or a different prior—to judge coin locations. Under the MAP hypothesis, subjects’ behavior would be independent of their prior (see Figure 1 and Table 1). This would also be the case if subjects merely neglected the prior learned in the estimation trials during the 2AFC trials and instead relied on a different and unchanged prior.

All trials began as described previously, with each block consisting 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 consisted of mostly estimation trials (375 estimation trials and 125 2AFC trials), whereas the second half consisted of mostly 2AFC trials (125 estimation trials and 375 2AFC trials). After the end of the first block, subjects took a brief (3–5 min) rest before beginning the second block with a different prior.

At the start of each day, subjects were instructed, from a prepared manuscript, on how to complete the estimation and 2AFC tasks. Subjects were told that someone behind them (the exact location not being...
important) was tossing coins, one at a time, into the “pond” (i.e., the screen). In the estimation trials, their task was to try to catch the coin by placing a net (i.e., 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 that the vertical component of their guess did not matter because the net spanned the whole height of the screen. For the 2AFC trials they would have to guess which of the unseen coins landed farther to the right. Instructions were provided on how to indicate their choice with a key press. To reduce the influence of uncontrolled cognitive strategies, subjects were also told that the person throwing the coins was not trying to help or hinder their progress or reacting to the choices they made.

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

### Data analysis

#### Estimation task

Independent of the decision-making strategy (sampling or MAP), our analysis assumes that subjects form beliefs by optimally combining priors and likelihoods. The combination of these two pieces of information gives rise to the well-known linear weighting of prior expectations and evidence (see Kording & Wolpert, 2004):

$$\hat{s} = (1 - r_{\text{reliance}})c + r_{\text{reliance}}s,$$

(23)

where $\hat{s}$ is the estimated coin location, $c$ is the cue, $\mu$ is the mean of the prior, and $r_{\text{reliance}}$ is the relative reliance on the likelihood, defined as $\sigma^2_{\text{sensory}}/(\sigma^2_{\text{sensory}} + \sigma^2_{\text{prior}})$. Intuitively, if the sensory information (likelihood) is very important relative to the prior (i.e., $\sigma^2_{\text{sensory}} \ll \sigma^2_{\text{prior}}$), then $r_{\text{reliance}} \approx 1$ and the expected stimulus position would be at the center of the sensory feedback/cue location. Similarly, if the prior is very important relative to sensory feedback (i.e., $\sigma^2_{\text{sensory}} \gg \sigma^2_{\text{prior}}$), then $r_{\text{reliance}} \approx 0$ and the estimated stimulus position would be at the prior’s mean. The reliance on the likelihood is an indirect measurement of the variance of a subject’s prior (Kording & Wolpert, 2004; Berniker et al., 2010; Vilares et al., 2012).

Our estimation experiment collected a set of experimentally defined cues, $c$, and the respective subjects’ estimation of the stimulus position, $s$. By using ordinary least squares estimation, it is possible to estimate $r_{\text{reliance}}$ without bias (Hastie, 2009). Similarly, we can manipulate Equation 23 to obtain the mean of the prior.

#### Estimating PSE and JND from psychometric data

We use a psychometric function (the cumulative Gaussian function) that matched the theoretical framework used in this study. In this way, we can directly compare the estimated quantities of our fits with the predicted quantities of the theoretical decision-making mechanisms. We model the probability of a decision as

$$p(z = 1|\delta) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\delta - \text{PSE}}{\sqrt{2}\text{JND}} \right) \right]$$

(24)

where $\delta$ is the discrepancy between stimuli and $z$ is the decision of the subject. We find the values of PSE and $\sigma_{\text{JND}}$ by a maximum-likelihood estimation algorithm.

The psychometric function in Equation 24 assumes that subjects make no distraction mistakes; given a sufficiently large discrepancy between the stimuli, 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 nonnegligible chance that subjects commit errors and respond randomly, independent of the discrepancy, with $\lambda$ probability. The psychometric curve can then be modified to accommodate this change as follows:

$$p(z = 1|s_1, s_2) = \frac{\lambda}{2} + (1 - \lambda)p(z = 1|\delta).$$

(25)

We find the parameters jointly by maximum-likelihood optimization. Every psychometric curve used in this study was adapted this way to include a lapse. For estimating the slope of PSE, JND, reliance in feedback, and mean of the prior, we dropped the first 500 trials for the first day and the first 100 trials for all other days. Confidence intervals and standard errors were computed using 1,000 bootstrap samples. Unless otherwise specified, all $t$ tests were two-sided and significance level was set to 0.05.

### Results

Although the 2AFC task is well studied, exactly which decision-making mechanism the brain uses to complete this task is not clear. For two popular classes of decision-making mechanisms, we mathematically derived the predicted psychometric curve and the influence of the prior on the JND. We then measured
the actual JND while we manipulated subjects’ priors in a behavioral 2AFC experiment. A comparison of the predictions with the experimental data was then used to examine which mechanism explains the data best and what the JND actually characterizes.

Decision-making mechanisms and the JND predictions

One of the central measures obtained from a psychometric curve is the JND, which quantifies how different two stimuli must be for subjects to reliably report them as distinct. Importantly, the JND is thought to be independent of individuals’ subjective beliefs or biases (Green, 1966). As a result, the JND is assumed to be a valuable measure of a subject’s sensory precision, regardless of what neural mechanisms are used to make decisions. We found that the JND can have very different interpretations depending on how the brain makes decisions, calling into question traditional experimental findings.

One prominent hypothesis for decision making is the MAP mechanism (Green, 1966). Under this hypothesis, subjects make their choice by comparing the most probable values of the two stimuli (Figure 1A). If the brain uses this mechanism, the resulting psychometric curve and JND are independent of a subject’s prior (Figure 1B, C). Under this condition the experimentally derived JND is directly related to a subject’s sensory noise (see Table 1). Most studies that use the 2AFC to measure sensory precision implicitly assume the MAP mechanism. Again, if this assumption is true, then the prior has no influence on the JND, making the approach particularly attractive. Alternative decision-making mechanisms, however, predict distinct results.

Another prominent hypothesis for decision making is the sampling mechanism (Ackley et al., 1985; Hoyer & Hyvärinen, 2003; Fiser, Berkes, Orbán, & Lengyel, 2010). Under this hypothesis, choices are made using approximations to the most probable stimulus values (Figure 1A). These approximations are computed by averaging \( k \) sample values drawn from each belief distribution (see Materials and methods). If the brain uses this mechanism, the JND does not measure sensory precision. Rather, the JND is influenced by sensory noise and prior beliefs (see Table 1 and Figure 1B, C). We also note the special case where \( k = 1 \) (Table 1), which is the matching hypothesis (Estes, 1950; Myers, 1976; Vulkan, 2000; Wozny et al., 2010).

The predictions from the sampling hypothesis are in stark contrast with the traditional interpretation of the 2AFC task. We find that when a subject’s prior is certain (relatively small \( \sigma_{\text{prior}} \)), the JND increases (see Figure 1C). Intuitively, we can interpret this as follows: As the prior becomes more and more certain, sensory information becomes less relevant, and distinguishing a difference between the two stimuli requires increasingly large differences. By investigating multiple variations of these decision-making hypotheses, we can derive a corresponding interpretation of the experimental JND. In the next section we present results from an experiment that exploits these differing interpretations.

Estimation trials measure subjective beliefs

Across 5 days, subjects performed two kinds of randomly interleaved trials (see Materials and methods). In the estimation trials subjects estimated the location of a hidden coin. In the 2AFC trials subjects had to decide which of two coins was farther to the right. To guess the coins’ locations, subjects were shown splashes (a likelihood) indicating where the coin landed (see Figure 2). Half the trials had a wide splash (providing poor evidence) and half the trials had a narrow splash. Similarly, the hidden coins were drawn from two distributions (the priors), one wide and the other narrow. Thus, on each day subjects were exposed to four conditions: hidden coins drawn from two priors (wide or narrow; WP or NP, respectively) with two likelihoods (wide or narrow; WL or NL, respectively). The four conditions were abbreviated as WP-WL, WP-NL, NP-WL, and NP-NL. Using these four conditions, we could monitor a subject’s subjective prior while simultaneously measuring their JND. By combining these results, we could quantify the relation between prior uncertainty and JND.

During the estimation trials we measured two important variables: the mean of the subjects’ prior and the linear relationship between the splash and the subject’s estimated coin location, which we refer to as the reliance on the likelihood (relative to the prior). The reliance on the likelihood is an indirect measurement of the variance of a subject’s prior (Kording & Wolpert, 2004; Berniker et al., 2010; Vilares et al., 2012). With these variables we could monitor both the mean and the variance of each subject’s subjective prior.

The true mean of the experimentally defined prior of the hidden coins (i.e., the middle of the screen) was the same for each day and condition. Typical subject means were very accurate and not significantly different from the correct mean, \( r(135) = 1.04, p = 0.3 \) (see Figure 3B for example subject). Pooling each subject’s data across days, we found differences across subjects, \( F(6, 18) = 3.26, p = 0.02 \), but not conditions, \( F(3, 18) = 0.06, p = 0.97 \) (two-way analysis of variance, or ANOVA). This suggested that overall the subjects learned the correct, condition-independent, experimentally defined mean.

For subjects to accurately estimate the hidden coin’s location, they required an accurate estimate of the
prior’s variability. Therefore, we examined the reliance on the likelihood (see Figure 3C for example subject). Using only the first 250 trials of the first day, we found significant differences across experimental priors, $F(1, 19) = 108.1, p < 0.01$, and likelihoods, $F(1, 19) = 21.74, p < 0.01$, but not subjects, $F(1, 19) = 0.874, p = 0.53$. This suggested that all subjects reacted to the four conditions early on, and did so similarly. Examining the reliance on the likelihood across 250-trial bins for the first day, we found that the distance from the optimal slope significantly diminished for the NP-WL condition, $t(26) = -2.44, p_{\text{one-sided}} = 0.01$, but not for the other conditions; NP-WL: $t(26) = 2.51, p_{\text{one-sided}} = 0.99$; WP-NL: $t(26) = -1.37, p_{\text{one-sided}} = 0.09$; WP-WL: $t(26) = -0.78, p_{\text{one-sided}} = 0.22$. Overall, these results suggest that subjects’ behavior converged quickly and that no significant learning was observable even during the first day.

The analyses determined that subjects learned quickly within the first day, but we also wished to know whether subjects’ behavior changed across days. Examining the last 250 trials of each day, we found significant differences across priors, $F(1, 123) = 324.8, p < 0.01$, likelihoods, $F(1, 123) = 91.1, p < 0.01$, and subjects, $F(6, 123) = 4.42, p < 0.01$, but not across days, $F(4, 123) = 1.43, p = 0.22$. Therefore, subjects’ overall responses did differ, but these differences were dominated by the changes across prior and likelihood conditions. Pooling the data across subjects and days, the overall reliance on likelihood was as follows [mean ± SE (optimal), $n = 34$]: NP-NL: 0.76 ± 0.014 (0.91); NP-WL: 0.47 ± 0.016 (0.39); WP-NL: 0.96 ± 0.007 (0.99); and WP-WL: 0.92 ± 0.006 (0.94). We note here that on the whole, both within and across subjects, these numbers are all statistically distinct from the theoretical optimum. Regardless, we are able to precisely characterize each subject’s prior. Furthermore, our analysis demonstrated that across conditions, the subjects had stable priors across days.

We conclude that subjects take into consideration the prior and the likelihood uncertainty when making a decision in the estimation task. Importantly, they learned a different prior for each imposed distribution of coins and we were able to measure this subjective belief. This will allow us to examine whether changes in subjects’ priors influenced their JNDs—a critical test for distinguishing the MAP and sampling hypotheses.

**2AFC trials quantify psychometric functions**

During the 2AFC trials, subjects were shown two splashes and asked to guess which coin landed farther to the right. As should be expected, the decision probabilities systematically varied with the disparity of the evidence (see Figure 4): The larger the distance between the two stimuli, the more certain the answer. From these data we can fit a psychometric curve and extract two important variables: the PSE and the JND. Again, this information is needed to test which decision-making mechanism the brain uses.

If subjects were biased (e.g., believing that the coin landed farther to the right), the PSE would quantify it. To test this, we estimated PSEs for each subject, condition, and day. We found that there was no significant bias in the PSE: $t(67) = 1.71, p = 0.091$, averaged across subjects, days, and conditions. By performing a three-way ANOVA, we found that there were no significant differences in PSE across priors, $F(1, 56) = 1.74, p = 0.19$, days, $F(4, 56) = 0.45, p = 0.73$, or subjects, $F(6, 56) = 2.25, p = 0.05$. Overall, these results suggest that subjects did not have strong biases in their decisions.

Next we measured the JNDs. The average JND was 0.086 ± 0.003 screen units (across subjects, days, and conditions); that is, subjects could reliably discriminate two coin splashes when they varied by roughly 8% of the screen, or 4.5 cm. A three-way ANOVA revealed that there were no significant differences in JNDs across subjects, $F(6, 56) = 1.38, p = 0.23$, or days, $F(4, 56) = 1.27, p = 0.28$. There were significant differences across priors, $F(1, 56) = 34.3, p < 0.001$, and the measured JNDS were larger in the wide prior condition (WP-WL; the data from the WP-NL trials are used in the analysis in the subsequent section). This suggests that subjects were self-consistent across days, but the prior did have some influence.

In summary, we were able to accurately measure individual JNDS and PSEs across conditions and days. We found that none of these estimates changed across days, a fact that we use later to pool data and test the relationship between reliance on the likelihood and a subjects’ JND.

**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. While it was clear that subjects were relying on their prior during the estimation trials, subjects may have been using a different strategy during the 2AFC trials. For example, subjects might simply use the centroids to choose between splashes (i.e., a maximum-likelihood estimate) and neglect their prior altogether. To eliminate this possibility, we examined the 2AFC trials in which the two coins had different splash sizes (WL and NL conditions). In such cases, the probability of a response is a function of both cue locations (not merely the difference between them) and the psychometric function is a surface (see Equation 22 in Materials and Methods...
methods and Figure 5A). If subjects use their prior, both sampling and MAP models predict that the PSE changes as a function of cue locations. Thus, these trials served as a control to establish that the prior was actually used in the 2AFC trials.

In agreement with the models, we found that the linear relationship between PSEs and cue locations changed with the condition’s prior. The PSE’s linear relationship for the NP condition was significantly smaller than that for the WP conditions within subjects.

Figure 3. Measuring subjects’ priors. (A) Data from the estimation trials (grouped by condition) for a typical subject on their first day. Axes are in normalized screen units. A linear regression to the data determines the subjective mean of the prior and the reliance on the likelihood. (B) Data grouped into 250-trial bins are used to display this subject’s estimated means during the first (left side) and second (right side) half of one day’s experiment. The collective averages for both conditions are displayed to the right of each panel. (C) The same data and groupings are used to display the reliance on the likelihood. Error bars for all plots are the standard error of the mean, and the correct theoretical values are displayed with dashed lines.

Figure 4. JND across days and two conditions. (A) Example data from a typical subject for two conditions and the resulting psychometric curves. The response probability indicates the proportion of trials in which subjects reported the second coin as being farther to the right. Red and blue dots are subject responses during the narrow and wide prior conditions, respectively. Note that the PSE is very close to zero; the subject did not have a bias in coin locations. Also note that the curves are very similar and thus have similar JNDs. (B) JNDs for each subject across days and two conditions (mean ± standard error of the mean).
paired $t$ test, $t(6) = -2.35$, $p = 0.028$ (Figure 5B). Thus, subjects used the prior during the 2AFC task.

**Maximizing, not sampling: Narrower priors do not give rise to larger JNDS**

By combining the results from the estimation and 2AFC trials we can directly test which decision-making mechanism subjects use. If the MAP hypothesis is true, JNDS should not vary with subject priors. However, if the sampling hypothesis is true, then having a narrower prior should give rise to larger JNDS. A regression analysis revealed no such trend. No single subject showed a significantly negative relationship between the reliance on likelihood and JND; $S1$: $t(8) = 3.19$, $p_{one-sided} = 0.99$; $S2$: $t(8) = 6.2$, $p_{one-sided} = 0.99$; $S3$: $t(8) = 3.5$, $p_{one-sided} = 0.99$; $S4$: $t(8) = 1.41$, $p_{one-sided} = 0.90$; $S5$: $t(8) = 1.23$, $p_{one-sided} = 0.87$; $S6$: $t(8) = 1.63$, $p_{one-sided} = 0.92$; $S7$: $t(6) = 1.6$, $p_{one-sided} = 0.92$ (Figure 6A). We performed the same analysis across subjects after pooling data across days. Again, we found that the data did not follow a negative trend: and there was no significant decrease in JNDS, $t(12) = 1.497$, $p_{one-sided} = 0.919$ (Figure 6B). Collectively, these results provide no evidence of the JND dependence on priors predicted by sampling.

To assess the overall ability of the MAP and sampling models to describe the subjects’ behavior, we again pooled data across days. For each model, we assumed that the across-subjects sensory noise (likelihood variance) was a free parameter. This is a sensible assumption since the subjects’ perception of the stimuli can only introduce further noise above and beyond that introduced by the experiment. The more noise in the sensory perception, the flatter the psychometric curve and the larger the JND. The sampling model had the additional free parameter, $k$, the sample size. Both model parameters were fit using the subject averages (Figure 6B, dotted and dashed lines). The fitted sensory noises for the MAP and sampling models were 35% and 31% larger than the true experimental standard deviation (0.05 screen units), respectively. This should be expected since the inherent noise in the visual system can only increase the uncertainty in the experimental stimuli. For the sampling model, the best fit for $k$ was $24$; 95% confidence interval [5, 222], bootstrapped across subjects and conditions. The sampling model’s large $k$ effectively approximated the MAP model. A quantitative comparison of the two models accounting for the difference in free parameters strongly favored the MAP model over the sampling model (Bayesian Information Criterion [BIC] = -62.6 and -47.4, respectively). As a further analysis, we also fit the two models to each individual subject. Here again the results favored the MAP over sampling when comparing BIC, $t_{paired}(6) = 18.61$, $p < 0.001$.

For multiple subjects, the measured JNDS increased with the reliance on the likelihood—a finding not predicted by either mechanism. Therefore, we performed a secondary, more conservative analysis that would not be affected by this effect. Using only the data obtained during the narrow prior condition (leftmost in Figure 6B), every subject’s average JND was below the matching prediction, $t(6) = -8.3$, $p < 0.001$, which means that the precision of their decisions was better.
than that possible with matching. That is, even if subjects had perfect visual acuity (i.e., their likelihood was equivalent to that experimentally imposed), these data rule out the matching mechanism. Similarly, we find sampling with $k = 2$ unlikely, $t(6) = -3.9, p = 0.006$. Matching as well as sampling with small $k$ seems incompatible with the data from this condition.

**Discussion**

The 2AFC paradigm is the technique of choice for measuring sensory uncertainty, widely interpreted to be independent of subjective priors. Here we have demonstrated that this interpretation depends crucially on how the brain makes decisions. By mathematically deriving the psychometric curves under two prominent decision-making mechanisms (MAP and sampling), we predicted how a subject’s prior would influence the experimentally obtained JND. To test these predictions we set up a visual task in which subjects either estimated or compared the locations of two stimuli. Our results strongly favor a MAP mechanism over sampling, supporting the common assumption that JNDs quantify sensory precision. Furthermore, we establish the 2AFC paradigm as a psychophysics tool for probing how the brain makes decisions.

While our results do not support sampling, we note that the two mechanisms become indistinguishable as the number of samples grows; that is, sampling with large $k$ will always appear like MAP. Our analysis shows that if subjects were sampling, they did so with a good number of samples. Limiting the stimulus exposure may limit the number of samples subjects can acquire. Sampling time was limited by the presentation of a visual mask that followed the stimulus after just 25 ms. It would thus seem that if subjects sample, they do so extremely rapidly. Future studies can address alternative techniques for limiting the potential number of samples in an effort to find further evidence for sampling. Regardless of the techniques employed, if the brain is making decisions with a large number of samples, distinguishing between MAP and sampling at the behavioral level will always remain a fundamental challenge.

An interesting finding was that subject JNDs actually increased rather than decreased with increasing reliance on likelihood. This was a finding that neither decision-making mechanism predicted. We can speculate that a subject’s likelihood may not be constant but rather may depend on many parameters (e.g., the proportion of the screen the subject needs to attend to, the size of the experimentally imposed noise, the prior distribution of hidden coins, a subject’s motivation, or some combination of these). Sources of uncertainty such as these would lead to increases in subjects’ uncertainty and measured JNDs without influencing the PSE, just as we’ve found. Regardless, precisely what caused this effect could be the subject of future studies.

At the behavioral level, there are theoretical and experimental studies that argue for sampling-based decisions. Research in decision making has demon-

![Figure 6. Comparing subject data and candidate decision-making theories. (A) Each subject’s measured JND plotted versus the reliance on the likelihood for each day (bootstrapped means). The gray dashed lines are the best linear fit to the change in JND across the reliance on the likelihood. Overlaid on each plot are the optimal MAP and sampling (for $k = 1, 2, 3$) predictions. (B) Each subject’s across-days averages are displayed. Error bars indicate 95% confidence intervals over the JND and the reliance on the likelihood. Dotted and dashed black lines display the best MAP and sampling ($k = 24$) models fit to all data, respectively.](http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933691/)

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933691/ on 06/06/2017
strated that the sampling mechanism is optimal under some circumstances, such as when one believes that the statistics of the environment are nonstationary or while the subject is actively learning (Sakai & Fukai, 2008a, 2008b; Vul et al., 2009). At the same time, work in cognitive (Gaijsmaier & Schoorler, 2008) and perceptual (Wozny et al., 2010; Battaglia, Kersten, & Schrater, 2011) tasks has found experimental evidence for sampling. These studies, however, employ tasks requiring multisensory cue combination and reinforcement learning, making them difficult to compare with our study. Nevertheless, we can speculate that the added steps in multimodal sensory integration and cognitive load could contribute to significant changes in decision-making strategies. Even more, sampling can be implemented in a number of ways. Yet we found that two prominent approaches are indistinguishable behaviorally (see Supplemental Appendix). Future work can address why some tasks appear to exhibit a sampling mechanism, whereas our 2AFC task favors a MAP mechanism.

Similarly, at the neural level, there are many theoretical and experimental studies suggesting that neurons encode information by sampling. Under these theories, spikes or firing rates from neurons or pools of neurons encode samples from the posterior distribution (Ackley et al., 1985; Hoyer & Hyvärinen, 2003; Fiser et al., 2010; Beck, Ma, Pitkow, Latham, & Pouget, 2012; Shadlen & Kiani, 2013). Consistent with these ideas, electrophysiological evidence of spontaneous neural activity has also been argued to support sampling (Berkes, Orbán, Lengyel, & Fiser, 2011; but see Okun et al., 2012). However, this evidence, from an observational study, lacks controlled manipulations of uncertainty and merits further work. Further studies will determine the degree to which sampling can help explain neural data. The data we present here are compatible with subjects sampling extremely rapidly. Neural data could be used to test this prediction.

There are also many reasons why the finding of a MAP mechanism is not surprising. Just as sampling is optimal under some circumstances, MAP is optimal under other circumstances (Kording, 2007; Bowman, Kording, & Gottfried, 2012). In fact, sampling is often viewed as an approximation to the MAP answer. Indeed, in our study, MAP is the optimal solution strategy. Many have speculated that since perceptual mechanisms are shared across species and are evolutionarily old, there was sufficient time and pressure to evolve statistically optimal strategies, namely MAP (Parker & Smith, 1990; Barkow, Cosmides, & Tooby, 1992; Weibel, Taylor, & Bolis, 1998).

Innumerable studies implicitly assume that the brain uses a MAP mechanism in the 2AFC paradigm (e.g., Ernst & Banks, 2002; Knill & Saunders, 2003; Alais & Burr, 2004; Stocker & Simoncelli, 2006; Fetsch et al., 2011; Girshick et al., 2011; Trommershäuser, Kording, & Landy, 2011). If this were not the case, the many studies that use psychometric curves to measure sensory precision would be called into question. Similarly, an accurate characterization of sensory precision is important for many phenomenological laws—for example, the law of comparative judgment, Weber’s law, and Fechner’s law (Green, 1966). Here we have attempted to validate this ubiquitous assumption for MAP. For lack of contrary evidence, psychophysics studies may continue to assume an MAP mechanism when implementing the 2AFC paradigm. However, the protocol we present here can be used for determining corrections to these phenomenological laws when a thorough examination of the brain’s decision-making mechanisms is needed.

**Keywords:** decision-making, psychophysics, Bayesian, sampling, maximum a posteriori, two-alternative forced choice (2AFC), just-noticeable difference (JND)

---

**Acknowledgments**

D. E. A. and K. P. K. are supported by PHS Program Project Grant NS44393. M. B. is supported by NIH Grant 1RO1NS063399. K. P. K. is supported by NS074044. H. L. F. is supported by Fundação para a Ciência e a Tecnologia (FGRHD/BD/33525/2008).

*Asterisk indicates that authors contributed equally.

Commercial relationships: none.

Keywords: decision-making, psychophysics, Bayesian, sampling, maximum a posteriori, two-alternative forced choice (2AFC), just-noticeable difference (JND)

**References**


Berniker, M., Voss, M., & Kording, K. (2010). Learning priors for Bayesian computations in the nervous system. PLoS One, 5(9), e12686.


