Appendix

Task instruction
Instructions were presented to participants on the computer screen. The exact instructions for participants on the first practice trial of the Experimental block were ‘Now let’s start Experiment 2! We are interested in how precise your memories are. Just like before, you will be presented with an arrow. Your task is to remember the color of the arrow. The arrow will disappear, and then you will see a wheel. Try to match the color as close as possible to the color presented. Move your mouse to the wheel and click the left mouse button to make a selection. Afterwards, a white line will appear. This is where you will report your confidence. If you are highly confident, click further from the wheel. Less confidence = click closer. Lastly, you will see feedback about your response. Click the left mouse button to begin’ and the exact instructions on the first experimental trial were ‘Now, experimental trials will begin. The directions are the same as the practice block. Now, the arrows will appear faster. Try to match the color of the arrow as close as possible to the color presented. Click the left mouse button to make your selection. Afterwards, report your confidence with the line. (far = more confident). Lastly, you will see feedback about your response. Click the left mouse button to begin’. On the surprise trial, participants were presented with this message ‘This is a surprise test. What was the direction of the arrow?’.

ZL_s model across all trials
The ZL_s model is the Zhang and Luck two-component mixture model (2008) that we modified to be fitted across participants, such that each participant contributes a single trial to the maximum likelihood estimation. The results are shown across all trials in Appendix Figure 1.

Relevant feature analysis
One possibility regarding the measured cost between pre- and post-surprise color reports is that, instead of getting suddenly worse at the surprise trial, participants’ performance was gradually getting worse over time, perhaps due to fatigue or boredom.
To test whether the data support this, we applied mixture models to the complete set of color-response errors, where we used a mixture model of two von Mises distributions with freely varying concentrations ($\kappa_1, \kappa_2$) and a third parameter determining the probability of a response being in the first von Mises ($P_1$). We allowed that the parameters of the model depend on the trial number. We used three versions of the model: In model 1, the parameters have no trial dependence at all; i.e., there is neither an effect of direction becoming relevant at trial 26 nor an effect of fatigue or practice. In model 2, the parameters $P_1, \kappa_1$ and $\kappa_2$ depend linearly on trial number (i.e. there are a constant term and a slope parameter for each of $P_1$, $\kappa_1$ and $\kappa_2$), but there is no jump at the surprise trial; this models the possibility of fatigue causing a trial dependence, without any effect of direction becoming relevant on trial 26. In model 3, we have one set of trial-independent parameters before the surprise trial, and a different set afterwards; this implements the hypothesis that a change of precision is caused solely by direction becoming relevant at the surprise trial.

We fitted each model by maximum likelihood estimation, with the results shown in appendix Table 1. As measured by the Akaike Information Criterion (to allow for the variable number of parameters between the models; AIC; Akaike, 1974), the model with the jump in the parameters at the surprise trial is strongly preferred over
model 2 with its linear dependence on trial, and even more strongly preferred over model 1 which has no trial dependence of any kind.

<table>
<thead>
<tr>
<th>Model</th>
<th>Trial dependence</th>
<th>LL</th>
<th># Parameters</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Constant</td>
<td>-818.1</td>
<td>3</td>
<td>1642.2</td>
</tr>
<tr>
<td>2</td>
<td>Linear</td>
<td>-783.8</td>
<td>6</td>
<td>1579.6</td>
</tr>
<tr>
<td>3</td>
<td>Jump at surprise trial</td>
<td>-775.7</td>
<td>6</td>
<td>1563.4</td>
</tr>
</tbody>
</table>

Appendix Table 1: Models and fit statistics for the relevant feature analysis.

Fixed effects comparison between pre and post-surprise color trials
We used a non-parametric model comparison process to further demonstrate the cost of making the direction feature relevant following the surprise trial. To do this, we compared the log likelihood (LL) of the null hypothesis, wherein sigma is yoked between pre and post-surprise color, to the alternative hypothesis, wherein sigma is allowed to vary between fits to the pre and post surprise data. First, we pooled the pre-surprise color data and the post-surprise color data across participants (i.e. fixed effects analysis). Next, the ZL_s model was fit to both sets of data with the parameters sigma and pU yoked between fits. The sum of the log likelihoods between fitted resulted in an LL of –862.9. Then, the alternative model was fit to the data, whereby sigma was allowed to vary between the pre-surprise and post-surprise data. This resulted in a LL of –822.3.

An LL difference of 40.6 already demonstrates the better fit of the alternative model. However, to assess the significance, a permutation test was conducted on the magnitude of the observed LL difference. First, we permuted membership between pre- and post-surprise trials. To these permuted data sets were then applied the same fitting procedures described for the observed data. In 1000 iterations, no LL difference between the models in the permuted data exceeded the observed difference, which can conservatively be stated as a p < .001 with regards to the preference of the alternative model to the null model.

Model fits across condition
In appendix Table 2 and Table 3, the ZL and ZL_s and VP model parameter fits to the different conditions are displayed, respectively; these are the fits described in the main text.

<table>
<thead>
<tr>
<th></th>
<th>σ</th>
<th>pU</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre surprise color</td>
<td>12.9(.26)</td>
<td>.015(.003)</td>
<td>-.03(.54)</td>
</tr>
<tr>
<td></td>
<td>( \bar{\theta} )</td>
<td>( \tau )</td>
<td>( \kappa _r )</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------</td>
<td>---------</td>
<td>----------------</td>
</tr>
<tr>
<td>Pre surprise color</td>
<td>29.62</td>
<td>12.30</td>
<td>214.5</td>
</tr>
<tr>
<td>Surprise direction</td>
<td>4.35</td>
<td>16.26</td>
<td>408.1</td>
</tr>
<tr>
<td>First post surprise direction</td>
<td>44.93</td>
<td>70.12</td>
<td>408.1</td>
</tr>
<tr>
<td>Post surprise direction</td>
<td>38.28</td>
<td>22.43</td>
<td>408.1</td>
</tr>
<tr>
<td>Post surprise color</td>
<td>17.70</td>
<td>9.91</td>
<td>214.5</td>
</tr>
</tbody>
</table>

Appendix Table 3: VP model parameter fits to the different conditions across all participants.

**Formulation of the VP model of van den Berg et al. (2012)**

**Ingredients**

The ingredients of the VP model are the von Mises distribution and the Gamma distribution. The von Mises distribution for an angle \( s \) is

\[
M(s - \mu; \kappa) = \frac{e^{\kappa \cos(s-\mu)}}{2\pi I_0(\kappa)}
\]  

[S.1]

with the parameters being the mean \( \mu \) and the concentration \( \kappa \). \( I_0(\kappa) \) is the associated Bessel function of order 0. Two other parameters may be used instead of \( \kappa \). One is the circular standard deviation \( \sigma \), given in degrees by

\[
\sigma = \frac{180}{\pi} \sqrt{-2 \ln \frac{I_1(\kappa)}{I_0(\kappa)}}
\]  

[S.2]

The other is the precision \( J \), defined as Fisher information, given by

\[
J = \frac{\kappa I_1(\kappa)}{I_0(\kappa)}
\]  

[S.3]

For high precision \( J \) or \( \kappa \) (or, equivalently, small \( \sigma \), i.e., for \( J \gg 1 \) the relations between these parameters are:

[S.4]
where, to quantify the error in the approximation, we use the standard mathematical notation “$O(1/\kappa)$” to mean a term that is at most of order $1/\kappa$. For low precision (or large $\sigma$)

$$J = \frac{\kappa^2}{2} (1 + O(\kappa^2)), \quad \sigma = \frac{180}{\pi} \sqrt{\frac{2}{\kappa}} + O(\kappa^2)$$  \[S.5]\]

The Gamma distribution with shape parameter $a$ and scale parameter $\tau$ is

$$G(J; a, \tau) = \frac{J^{a-1} e^{-J/\tau}}{\tau^a \Gamma(a)}$$  \[S.6]\]

where $\Gamma$ is the gamma function. The VP model uses it for the distribution of $J$ in a mixture of von Mises distributions. The mean value of $J$ is $\bar{J} = a\tau$, and the standard deviation is $\sqrt{\alpha} = J/\sqrt{\alpha}$. For large $a$, the distribution is narrowly concentrated around the mean. For small $a$, the distribution is broader and skewed towards low $J$. At large $J$, the distribution falls off approximately exponentially, but at small $J$ the distribution diverges if $a < 1$, becoming non-integrable when $a \rightarrow 0$.

**The VP model**

Let there be a memory of an original stimulus at angle $s$. In the VP model, the basic response from the memory system is given by a value $\hat{s}$ (an estimate of $s$) given by a distribution that is a continuous mixture of von Mises distributions, with $J$ given by a Gamma distribution:

$$p_{\text{basic}}(\hat{s} | s) = \int_{0}^{\infty} dJ \ G(J; a, \tau) \ M(\hat{s} - s; \kappa(J)) \quad \text{[S.7]}$$

The actual measured response $r$ is given by convoluting this with a distribution of response noise that is a von Mises distribution with concentration parameter $\kappa_r$. Thus

$$p(r | s) = \int_{0}^{2\pi} d\hat{s} \ M(r - \hat{s}; \kappa_r) \int_{0}^{\infty} dJ \ G(J; a, \tau) \ M(\hat{s} - s; \kappa(J)) \quad \text{[S.8]}$$

where $\kappa(J)$ is given by the inverse of the functional relationship given in Eq. (S.3).
Van den Berg et al. (2012) use \( J = a \tau \) and \( \tau \) as independent parameters rather than \( a \) and \( \tau \). They also give \( J \) a power-law dependence on set size, which we do not need in the present paper. Note that the exponent \( \alpha \) for that set-size dependence is not related to the parameter \( a \) that we use here for the shape parameter.

References