SUPPLEMENTARY INFORMATION

A resource-rational theory of set size effects in visual working memory

Ronald van den Berg & Wei Ji Ma

Contents

Relation between Fisher information *J* and concentration parameter κ

Since we are only considering stimuli with circular domains, we assume that memory encoding errors follow a Von Mises distribution with a concentration parameter κ,

$$p(\varepsilon \mid \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\varepsilon)}, \tag{S.1}$$

where I_0 is the modified Bessel function of the first kind of order 0. We measure encoding precision as Fisher information, *J*, which measures the performance of the best possible unbiased decoder. Substituting Eq. (S.1) into the definition of Fisher information, we find that *J* and κ are one-to-one related through

$$J = \kappa \frac{I_1(\kappa)}{I_0(\kappa)}. \tag{S.2}$$

Encoding precision *J* is a monotonically increasing function of κ and therefore invertible. However, the inverse is not analytic, so we will use numerical inversion to compute the mapping from *J* to κ when fitting models.

Theoretical proofs of some properties of the resource-rational model for local tasks

In this section, we prove three properties of the general model that we presented for "local" tasks, i.e., tasks in which responses depend on a single item. This model is characterized by Eq. (11) in the main text,

$$\bar{J}_{\text{optimal},i}\left(p_i;\lambda,\tau\right)=\underset{\bar{J}}{\operatorname{argmin}}\left(p_i\bar{c}_{\text{behavioral}}\left(\bar{J};\tau\right)+\lambda\bar{J}\right), \tag{S3}$$

where $\bar{J}\geq 0$, $p_i\in\left[0,1\right]$, and $\lambda\geq 0$. We will also use the derivative of the expected total cost,

$$\bar{c}_{\text{total}}'\left(\bar{J}\right)=p_i\bar{c}_{\text{behavioral}}'\left(\bar{J}\right)+1,$$

where $\bar{c}_{\text{behavioral}}'$ is the derivative of the expected behavioral cost.

We will now prove that the following three claims hold under very general assumptions about the shape of the expected behavioral cost function in this model:

Claim 1. When neural coding is costly ($\lambda>0$), it is optimal to encode items with a finite amount of resource;

Claim 2. It is sometimes optimal to not encode a task-relevant item;

Claim 3. When each item is equally likely to be probed, $p_i=1/N$, the optimal amount of resource per item decreases with set size.

Assumptions about the expected behavioral cost

We construct our proofs under two intuitive and general assumptions about the expected behavioral cost function $\bar{c}_{\text{behavioral}}\left(\bar{J}\right)$:

Assumption 1. Expected behavioral cost is a monotonically decreasing function of resource: whenever more resource is invested, the expected behavioral cost is lower. This means that $\bar{c}_{\text{behavioral}}'\left(\bar{J}\right)\leq 0$ for all $\bar{J}$.

Assumption 2. A law of diminishing returns (Mankiw, 2004): when adding a bit of extra resource, the resulting decrease in $\bar{c}_{\text{behavioral}}\left(\bar{J}\right)$ is lower in magnitude when $\bar{J}$ is higher. This means that $\bar{c}_{\text{behavioral}}'\left(\bar{J}\right)$ is monotonically increasing, i.e. $\bar{c}_{\text{behavioral}}''\left(\bar{J}\right)\geq 0$ for all $\bar{J}$. As a consequence, $\bar{c}_{\text{behavioral}}'\left(\bar{J}\right)$ takes its lowest value at $\bar{J}=0$ and its largest as $\bar{J}\to\infty$.

Both assumptions are satisfied by the behavioral cost function that we used for fitting human data, namely $c_{\text{behavioral}}(\varepsilon;\beta)=|\varepsilon|^{\beta}$. Examples of the expected behavioral cost function under this choice and its first and second derivative are presented in Fig. S2.

Three scenarios

We now return to the problem of calculating $\bar{J}_{\text{optimal}}$, Eq. (S3). We are interested in the value $\bar{J}\in[0,\infty)$ that minimizes the expected total cost, $\bar{c}_{\text{total}}\left(\bar{J}\right)$. We separately consider the following

three scenarios: the minimum lies on the left boundary (0), on the right boundary (∞), or in between.

Scenario 1: $\bar{c}_{\text{total}}\left(\bar{J}\right)$ *is monotonically decreasing across the domain of* $\bar{J}$*, so* $\bar{J}_{\text{optimal}} \to \infty$.

When does this happen? The monotonic decrease means that $\bar{c}_{\text{total}}'\left(\bar{J}\right) \leq 0$ for all $\bar{J}$, or equivalently, $p_i \bar{c}_{\text{behavioral}}'\left(\bar{J}\right) \leq -1$ for all $\bar{J}$. Since we assume $\bar{c}_{\text{behavioral}}'\left(\bar{J}\right)$ to be monotonically increasing (Assumption 2), its largest value is attained at $\bar{J} \to \infty$. Therefore, $p_i \bar{c}_{\text{behavioral}}'\left(\bar{J}\right) \leq -1$ is equivalent to $p_i \bar{c}_{\text{behavioral}}'\left(\infty\right) \leq -1$, or (using Assumption 1) $p_i \left|\bar{c}_{\text{behavioral}}'\left(\infty\right)\right| \geq \lambda$. This means that it is optimal to invest infinite resource when p_i exceeds a critical value p_∞:

$$\bar{J}_{\text{optimal}} = \infty \text{ when } p_i \geq p_\infty \equiv \frac{\lambda}{\left|\bar{c}_{\text{behavioral}}'\left(\infty\right)\right|}.$$

The condition $p_i \geq p_\infty$ is satisfied when λ=0. This makes sense: when neural cost plays no role, there is no reason to not invest more. Other than that, the condition will rarely if ever be satisfied, since every expected behavioral cost function that we can think of has the property $\left|\bar{c}_{\text{behavioral}}'\left(\infty\right)\right| = 0$: as the amount of invested resource approaches infinity, there is no behavioral benefit in investing more resource (note that p_∞ has a domain [0,∞), not [0,1]). Therefore, unless neural cost plays no role, we do not expect it to be optimal to invest an infinite amount of resource in an item.

In tasks where p_i is one-to-one related to set size, the above result can be reformulated in terms of set size. In particular, when probing probabilities are equal, $p_i = \frac{1}{N}$, the above result implies that there exists a set size N_∞ (in general not an integer) below which it is optimal to invest infinite resource in each item:

$$\bar{J}_{\text{optimal}} = \infty \text{ when } N \leq N_\infty \equiv \frac{1}{p_\infty} = \frac{\left|\bar{c}_{\text{behavioral}}'\left(\infty\right)\right|}{\lambda}.$$

Scenario 2: $\bar{c}_{\text{total}}\left(\bar{J}\right)$ *is monotonically increasing across the domain of* $\bar{J}$*, so* $\bar{J}_{\text{optimal}} = 0$.

The monotonic increase means that $\bar{c}_{\text{total}}'(\bar{J}) \geq 0$ for all $\bar{J}$, or equivalently, $p_i \bar{c}_{\text{behavioral}}'(\bar{J}) \geq -\lambda$ for all $\bar{J}$. Since we assume $\bar{c}_{\text{behavioral}}'(\bar{J})$ to be monotonically increasing (Assumption 2), its smallest value is attained at $\bar{J} \to \infty$. Therefore, $p_i \bar{c}_{\text{behavioral}}'(\bar{J}) \geq -\lambda$ is equivalent to $p_i \bar{c}_{\text{behavioral}}'(0) \geq -\lambda$, or (using Assumption 1) $p_i \left| \bar{c}_{\text{behavioral}}'(0) \right| \leq \lambda$. This means that it is optimal to invest no resource when p_i is smaller than or equal to a critical value p_0:

$$\bar{J}_{\text{optimal}} = 0 \text{ when } p_i \leq p_0 \equiv \frac{\lambda}{\left| \bar{c}_{\text{behavioral}}'(0) \right|}.$$

A similar condition was derived in our earlier work (De Silva & Ma, 2018) for the case of a fixed total amount of resource (hard constraint).

The condition $p_i \leq p_0$ is satisfied when p_i=0. This makes sense: when an item never gets probed, one should not invest any resource. More generally, when probing probability is sufficiently low, the behavioral cost function is sufficiently shallow at 0, and neural cost is sufficiently important, it is not worth it to invest any resource on encoding. The expression for p_0 also makes clear that the optimal amount of resource is never 0 when the slope of the behavioral cost function at 0 approaches $-\infty$.

In tasks where p_i is one-to-one related to set size, the above result can be reformulated in terms of set size. In particular, when probing probabilities are equal, $p_i = \frac{1}{N}$, the above result implies that there exists a set size N_0 (in general not an integer) beyond which it is optimal to not invest any resource in any item:

$$\bar{J}_{\text{optimal}} = 0 \text{ when } N \geq N_0 \equiv \frac{1}{p_0} = \frac{\left| \bar{c}_{\text{behavioral}}'(0) \right|}{\lambda}.$$

Intuitively, this means that when set size is too large, the chances of success are too low and one should not even try.

Scenario 3: $\bar{c}_{\text{total}}(\bar{J})$ has a stationary point, so $\bar{J}_{\text{optimal}}$ is finite and nonzero.

We will now consider the remaining scenario, which is the complement of Scenarios 1 and 2; in particular, we can take λ>0 and p_i>0. The stationary point of $\bar{c}_{\text{total}}(\bar{J})$ will always be a minimum,

since the second derivative $\bar{c}_{\text{total}}''(\bar{J})$ is equal to $\bar{c}_{\text{behavioral}}''(\bar{J})$, which is always positive (Assumption 2). At the minimum, we have $\bar{c}_{\text{total}}'(\bar{J}) = 0$, from which it follows that $\bar{c}_{\text{behavioral}}'(\bar{J}) = -\frac{1}{p_i}$ at the minimum. Since the left-hand side is monotonically increasing as a function of $\bar{J}$ (Assumption 2), the minimum is either a single point or a single interval, but there cannot be multiple disjoint minima. Graphically, this equation describes the intersection between $\bar{c}_{\text{behavioral}}'(\bar{J})$, which is a monotonically increasing function, and a flat line at a value $-\frac{\lambda}{p_i}$ (Fig. S3). The value of $\bar{J}$ at which this intersection occurs necessarily increases with p_i.

Three regimes for probing probability

So far, we have assumed a given probing probability p_i. Now suppose that for a given $\bar{c}_{\text{behavioral}}(\bar{J})$ and a given λ, we increase p_i from 0 to 1:

- The first regime is $p_i \le p_0$. There, Scenario 2 applies and $\bar{J}_{\text{optimal}} = 0$: the item does not get encoded at all.
- The second regime is $p_0 < p_i < p_\infty$; there, Scenario 3 applies and $\bar{J}_{\text{optimal}}$ monotonically increases with p_i.
- The third regime is $p_i \ge p_\infty$. There, Scenario 1 applies and $\bar{J}_{\text{optimal}} = \infty$: the item gets encoded with infinite resource.

Even though not each regime might exist, the model predicts across regimes that $\bar{J}_{\text{optimal}}$ increases monotonically with p_i (Fig. 1F in main text).

Three regimes for set size

We can similarly examine the experimentally important special case of equal probing probabilities, $p_i = \frac{1}{N}$:

- The first regime is $N \le N_\infty$. There, Scenario 1 applies and $\bar{J}_{\text{optimal}} = \infty$: all items are encoded with infinite resource.
- The second regime is $N_\infty < N < N_0$. There, Scenario 3 applies and $\bar{J}_{\text{optimal}}$ monotonically decreases with N.
- The third regime is $N \ge N_0$. There, Scenario 2 applies and $\bar{J}_{\text{optimal}} = 0$: no items are encoded at all.

Even though not each regime might exist, the model predicts across regimes that $\bar{J}_{\text{optimal}}$ decreases monotonically with N (Fig. 1G in main text).

Conclusion

In conclusion, given Eq. (S3) and two additional assumptions, we have proven the following:

- Investing infinite resource in an item is only optimal when $p_i \left|\overline{c}_{\text{behavioral}}'(\infty)\right| \geq \lambda$. In practice, this might only happen when neural cost is unimportant (λ=0). This proves Claim 1.
- Investing no resource in an item is optimal when $p_i \left|\overline{c}_{\text{behavioral}}'(0)\right| \leq 1$. This can happen even when the probing probability p_i is nonzero. This proves Claim 2.
- $\overline{J}_{\text{optimal}}$ is a monotonically increasing function of p_i. In particular, if $p_i = \frac{1}{N}$, then $\overline{J}_{\text{optimal}}$ is a monotonically decreasing function of N. This proves Claim 3.

All three results hold more generally than we have shown here: we can replace the neural cost term $1/\overline{J}$ in Eq. (S3) by any function $c_{\text{neural}}\left(\overline{J}\right)$ whose derivative is positive and monotonically increasing. The proofs proceed along the same lines.

Special case: fixed-precision model

For the fixed-precision model (variable-precision model with $\tau \downarrow 0$), Eqs. (9) and (10) in the main text combine to take the form

$$\overline{c}_{\text{behavioral}}(J) = \int_{-\pi}^{\pi} c_{\text{behavioral}}(\varepsilon)\,\text{VM}(\varepsilon; J)\,d\varepsilon.$$

We wish to evaluate $p_0 \equiv \frac{\lambda}{\left|\overline{c}_{\text{behavioral}}'(0)\right|}$. First, we evaluate the derivative of $\overline{c}_{\text{behavioral}}(J)$ using the chain rule:

$$\frac{d\overline{c}_{\text{behavioral}}}{dJ} = \frac{d\overline{c}_{\text{behavioral}}}{dk}\frac{dk}{dJ}. \tag{S4}$$

Using Eq. (S.1), the first factor is

$$\begin{aligned}\frac{d\overline{c}_{\text{behavioral}}}{dk} &= \frac{d}{dk}\int_{-p}^{p} c_{\text{behavioral}}(\theta)\frac{1}{2\pi I_0(\kappa)}e^{\kappa\cos(\theta)}d\theta \\ &= \frac{1}{2\pi}\int_{-p}^{p} c_{\text{behavioral}}(\theta)\frac{d}{dk}\left(\frac{1}{I_0(\kappa)}e^{\kappa\cos(\theta)}\right)d\theta \\ &= \frac{1}{2\pi}\int_{-p}^{p} c_{\text{behavioral}}(\theta)\left(-\frac{I_0'(\kappa)}{I_0(\kappa)^2}e^{\kappa\cos(\theta)}+\frac{\cos(\theta)}{I_0(\kappa)}e^{\kappa}\right)d\theta \\ &= \frac{1}{2\pi}\int_{-p}^{p} c_{\text{behavioral}}(\theta)\left(-\frac{I_1(\kappa)}{I_0(\kappa)^2}e^{\kappa\cos(\theta)}+\frac{\cos(\theta)}{I_0(\kappa)}e^{\kappa}\right)d\theta.\end{aligned} \tag{S5}$$

where in the last line we used $I_0'(\kappa) = I_1(\kappa)$ (see Eq. 9.6.27 in (Abramowitz & Stegun, 1972)).

We next evaluate the second factor in Eq. (S4) using Eq. (S.2):

$$\frac{d\kappa}{dJ} = \left(\frac{dJ}{d\kappa}\right)^{-1} = \left(\frac{d}{d\kappa}\left(\frac{\kappa I_1(\kappa)}{I_0(\kappa)}\right)\right)^{-1} = \left(\frac{\kappa I_0(\kappa)^2 - \kappa I_1(\kappa)I_0'(\kappa)}{I_0(\kappa)^2}\right)^{-1} = \left(\kappa\left(1-\frac{I_1(\kappa)^2}{I_0(\kappa)^2}\right)\right)^{-1}, \tag{S6}$$

where in the third equality, we used $\frac{d}{d\kappa}(\kappa I_1(\kappa)) = \kappa I_0(\kappa)$ (see Eq. 9.6.28 in (Abramowitz & Stegun, 1972)). We now combine Eqs. (S5) and (S6) into Eq. (S4), and the result in turn in the expression for p_0. We also realize that the limti $J \downarrow 0$ is, using Eq. (S.2), equivalent to the limit $\kappa \to 0$. Putting everything together, we find

$$p_0 = 1/\lim_{\kappa\downarrow 0}\left|\frac{\kappa\left(1-\frac{I_1(\kappa)^2}{I_0(\kappa)^2}\right)}{\frac{1}{2\pi}\int_{-p}^{p} c_{\text{behavioral}}(\theta)\left(-\frac{I_1(\kappa)}{I_0(\kappa)^2}e^{\kappa\cos\theta}+\frac{\cos\theta}{I_0(\kappa)}e^{\kappa}\right)d\theta}\right| = 0.$$

We conclude that in our theory for delayed estimation, assuming the expected behavioral cost function from the fixed-precision model, it is only optimal to invest no resource at all into an item when that item has zero probability of being probed.

Optimal decision rule for the change detection task

In our simulation of the change detection task, we assume that observers use a Bayesian decision rule. This rule is to report "change" whenever the posterior ratio of change presence over change absence exceeds 1,

$$\frac{p\left(\text{change present}\,|\,\mathbf{x},\mathbf{y}\right)}{p\left(\text{change absent}\,|\,\mathbf{x},\mathbf{y}\right)} > 1,$$

where **x** and **y** denote the vectors of noisy measurements of the items in the first and second displays, respectively. Under the Von Mises noise assumption, and assuming a flat prior on change presence, this decision rule evaluates to (Keshvari, van den Berg, & Ma, 2013)

$$\frac{1}{N}\sum_{i=1}^{N}\frac{I_0\left(\kappa_{\mathrm{x},i}\right)I_0\left(\kappa_{\mathrm{y},i}\right)}{I_0\left(\sqrt{\kappa_{\mathrm{x},i}^2+\kappa_{\mathrm{x},i}^2+2\kappa_{\mathrm{x},i}\kappa_{\mathrm{y},i}\cos\left(y_i-x_i\right)}\right)} > 1,$$

where $\kappa_{\mathrm{x},i}$ and $\kappa_{\mathrm{y},i}$ denote the concentration parameters of the Von Mises distributions associated with the observations of the items at the i^{th} location in the first and second displays, respectively. The predicted probability of a correct response for a given resource vector, $p\left(\text{error}|\bar{\mathbf{J}}\right)$, is not analytic, but can easily be computed using Monte Carlo simulations.

Supplementary Tables

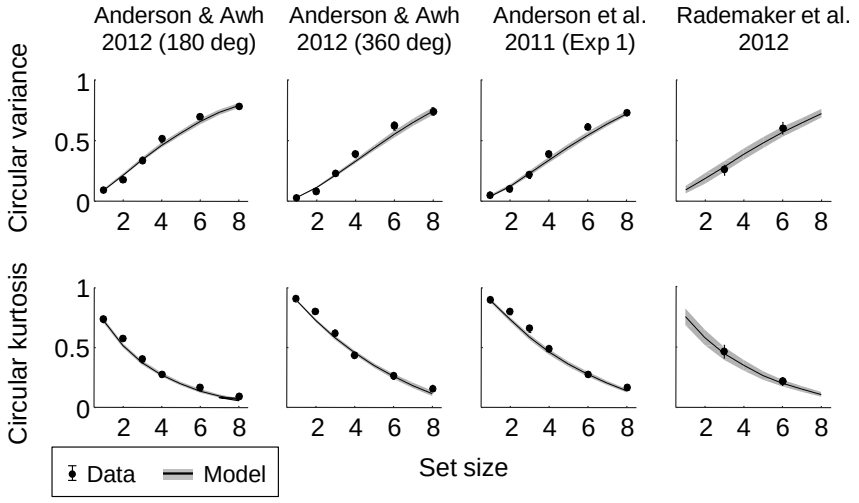
Table S1. Subject-averaged parameter estimates of the resource-rational model fitted to data from 11 previously published experiments. See Table 1 in main text for details about the experiments and references to the papers in which the experiments were originally published.

Experiment	λ	τ	β
E1	$(4.48 \pm 0.66)\cdot 10^{-3}$	16.7±2.1	1.97±0.43
E2	$(2.66 \pm 0.28)\cdot 10^{-3}$	12.6±1.2	$(7.1\pm1.2)\cdot 10^{-2}$
E3	$(3.47 \pm 0.34)\cdot 10^{-3}$	17.9±2.1	0.228±0.040
E4	$(4.5 \pm 1.0)\cdot 10^{-3}$	7.3±1.7	0.176±0.064
E5	$(4.15 \pm 0.37)\cdot 10^{-3}$	18.5±3.0	0.426±0.089
E6	$(4.89 \pm 0.57)\cdot 10^{-3}$	7.6±1.4	0.66±0.14
E7	$(6.1 \pm 1.2)\cdot 10^{-3}$	5.7±1.6	1.64±0.90
E8	$(2.45 \pm 0.37)\cdot 10^{-3}$	28.8±7.0	1.53±0.67
E9	$(3.10 \pm 0.50)\cdot 10^{-3}$	23.8±2.4	1.52±0.63

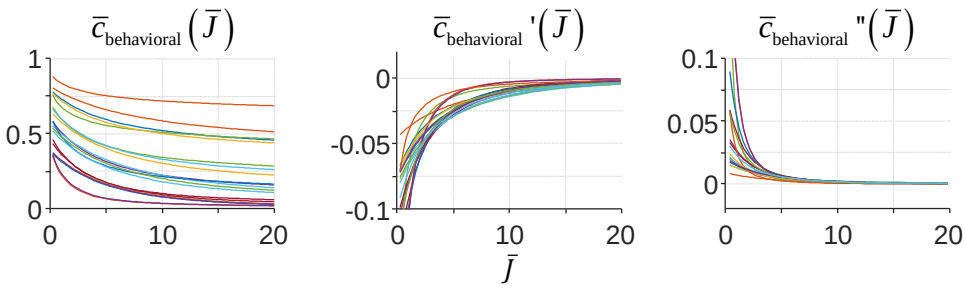
Table S2. Overview of AIC-based model comparison results reported in the main text, together with results from corresponding comparisons with five-fold cross validation (see Methods in main text). Each comparison is between the main version of the resource-rational model, Eq. (13), and the model listed in the first column of the table. Negative AIC differences and positive cross-validation log likelihood differences indicate an advantage of the resource-rational model over the alternative model. In all comparisons, these differences have opposite signs, which means that the AIC-based results are consistent with the cross-validation results.

Model with which the main model is compared	**AIC difference**	**Cross-validation log likelihood difference**
Descriptive power-law model	−5.27±0.70	2.59±0.39
Descriptive fixed-resource model	−13.9±1.4	8.4±1.0
Descriptive unconstrained model	3.49±0.93	−1.26±0.49
Rational model variant: equal precision	−110±10	56±4.7
Rational model variant: $c_{\text{behavioral}}=\lvert\varepsilon\rvert$	−14±2.8	7.1±1.4
Rational model variant: $c_{\text{behavioral}}=\varepsilon^2$	−24.4±4.1	12.2±2.0
Rational model variant: $c_{\text{behavioral}}=-\cos(\varepsilon)$	−19.5±3.5	9.8±1.8
Rational model variant: $c_{\text{behavioral}}$ as in Sims (2015)	−5.3±1.8	4.7±0.74

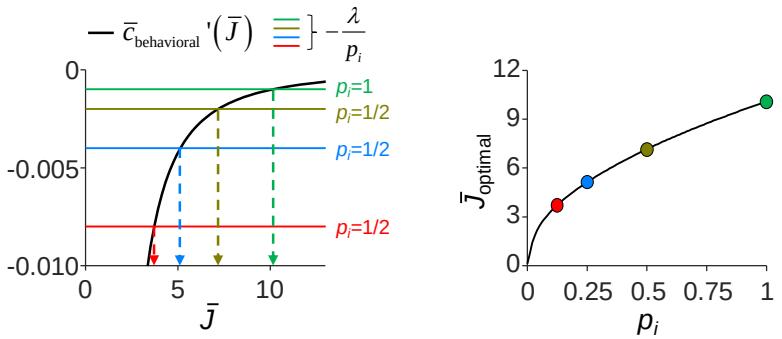
Supplementary Figures



Supplementary figure S1. Fits to the three delayed-estimation benchmark data sets that were excluded from the main analyses. Circular variance (top) and circular kurtosis (bottom) of the estimation error distributions as a function of set size, split by experiment. Error bars and shaded areas represent 1 s.e.m. of the mean across subjects. The first three datasets were excluded from the main analyses on the ground that they were published in papers that were later retracted (Anderson & Awh, 2012; Anderson et al. 2011). The Rademaker et al. (2012) dataset was excluded from the main analyses because it contains only two set sizes, which makes it less suitable for a fine-grained study of the relationship between encoding precision and set size.



Supplementary figure S2. Examples of the expected behavioral cost function and its first and second derivative under a behavioral cost function $c_{\text{behavioral}}(\varepsilon)=|\varepsilon|^2$. Different colors represent different choices of parameters β and τ (randomly drawn).



Supplementary figure S3. Graphical illustration of the solution to the cost-minimization problem that determines the value of $\bar{J}_{\text{optimal}}$. The cost-minimizing value of solution of $\bar{J}$ lies at the intersection between the derivative of the expected behavioral cost function (black curve) and a flat line at a value $-\lambda/p_i$ (colored lines). This value (indicated with arrows) necessarily increases with p_i. The parameter values used in this simulation were the same as in Fig. 1C of the main text.

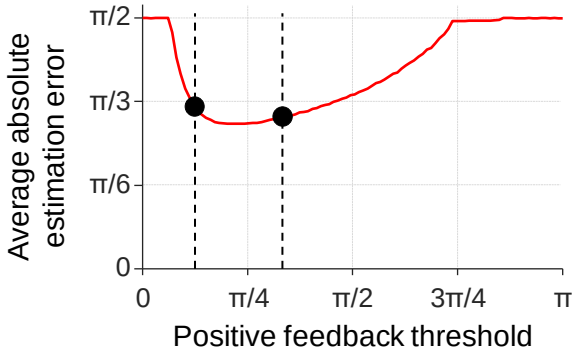


Figure S4. Model predictions for a delayed-estimation task with binary feedback (*N*=5). In this experiment, subjects receive positive feedback (e.g., "correct") when their estimation error is smaller than the positive feedback threshold and negative feedback (e.g., "error") otherwise. We modelled this using a behavioral cost function that maps errors below the feedback threshold to a cost of 0 and errors larger than this threshold to 1. The model predicts that subjects do not invest any resource when the threshold is very small or very large, such that the expected absolute estimation error is $\pi/2$ (guessing). This makes intuitive sense: for very small thresholds, the task is too difficult and for very large thresholds it is too easy to justify investing resource. In an intermediate regime, the prediction is U-shaped and contains a region in which the predicted estimation error barely changes as a function of feedback threshold. In this region, any performance benefit from increasing the amount of invested resource is almost exactly outdone by the added neural cost. The dashed lines show the feedback thresholds corresponding to the "high precision" and "low precision" conditions in the experiment by Nassar et al. (2017). Under the chosen parameter settings (λ=0.008, τ=30), our model predicts that the average absolute estimation errors in these two conditions (black circles) are very similar to each other.

Supplementary References

Abramowitz, M., & Stegun, I. (1972). *Handbook of mathematical functions*. *Handbook of Mathematical*.

Anderson, D. E., & Awh, E. (2012). The plateau in mnemonic resolution across large set sizes indicates discrete resource limits in visual working memory. *Attention, Perception & Psychophysics*, *74*(5), 891–910. https://doi.org/10.3758/s13414-012-0292-1

Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *The Journal of Neuroscience*, *31*(3), 1128–38. https://doi.org/10.1523/JNEUROSCI.4125-10.2011

De Silva, N., & Ma, W. J. (2018). Optimal allocation of attentional resource to multiple items with unequal relevance. *BioRxiv*.

Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No Evidence for an Item Limit in Change Detection. *PLoS Computational Biology*, *9*(2).

Mankiw, N. G. (2004). *Principles of economics. Book* (Vol. 328). https://doi.org/10.1017/CBO9780511511455

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*(13), 21. https://doi.org/10.1167/12.13.21