Supplemental Material

Subjects and Apparatus

Four White Carneaux pigeons (Palmetto Pigeon Plant, Sumter SC), 5-10 years old at the beginning of this experiment, were the subjects. Testing was conducted 5-7 days a week. The pigeons were maintained at approximately 85% of their free-feeding weights. Reinforcement for correct *same/different* choices was 2.5-5.0 seconds access to a lighted hopper filled with mixed grain. The pigeons had free access to grit and water and a 14-hr light, 10-hr dark cycle in the colony room where they were housed individually.

Pigeons were tested in two identical (35.9-cm wide x 45.7-cm deep x 51.4-cm high) custom built wooden test chambers and grain hoppers. Pecks to a computer monitor (40.3-cm color monitor, 800 x 600 pixel resolution, Eizo T550, Ishikawa, Japan) were detected by an infrared touchscreen (17" Unitouch, Carroll Touch, Round Rock, TX).

Stimulus displays consisted of two color pictures (each 5.7 x 3.8 cm of visual angles 68.7° H X 72.9° W) and a white rectangle (2.5 x 2.4 cm) on a black background. The two pictures were vertically aligned with a 1.28-cm gap between them. The top picture was centered 20.63 cm from the left edge and 18.75 cm from the top of the front panel. The bottom of the white rectangle was horizontally aligned with the bottom of the lower picture with a 1.4-cm gap between them and located to the right side of the lower picture.

Statistical Analyses of Baseline Performance

A three-way repeated measures ANOVA of Testing (1, 2), x Delay (1s, 10s) x Trial Type (*same/different*) conducted on baseline percent correct yielded no main effects or interactions (all Fs < 7.2, $p > .076$]). Mean percent correct for the 1-s delay *same* trials was 87.9 ± 2.4 and for *different* trials was 84.7 ± 3.5. Mean percent correct for the 10-s delay *same* trials was 85.9 ± 4.3 and for *different* trials was 78.6 ± 5.1. Only those *different* trial performances were used in analyses, modeling, and for the No-PI condition in figures.

Elapsed Times

If the interfering stimulus is presented m trials before the current trial, the time between the offset of the interfering stimulus and the onset of the current test stimulus is

$$T_I = m \cdot (\text{sample viewing time} + \text{ITI} + \text{response time} + \text{reward time}) + (m+1) \cdot \text{retention delay time}$$
.

The mean time (mean of the individual median times) to complete 20 responses to the sample stimulus was 23.8 ± 3.2 s in the 1-s delay condition and 33.3 ± 2.0 s in the 10-s delay condition. Response time was 1.44 ± 0.14 s in the 1-s delay condition and 1.53 ± 0.15 s in the 10-s delay condition. Reward (hopper) time was the time for which reward was presented multiplied by the mean proportion correct; it was 2.60 ± 0.51 s in the 1-s delay condition and 2.15 ± 0.51 s in the 10-s delay condition. We used the times for each individual subject in the fit to that subject.

Signal Detection Theory Model of Proactive Interference

We use signal detection theory to model this task. The basic idea is that a subject decides "same" or "different" based on its belief that the current sample stimulus followed the interfering sample stimulus ("different") or preceded it ("same"). To determine this belief, the subject compares the elapsed time since the offset of the current sample to the elapsed time since the offset of the interfering sample. We assume that the internal representation of elapsed time, denoted *t*, is noisy and can be described by a Weber-Fechner law:

$$t \sim \text{Normal}\left(a \log T, \sigma^2\right),$$

where *T* is the true elapsed time, *a* is a constant, and σ is the standard deviation of the noise. Specifically, for the current and interfering samples, indicated by an indices *C* and *I* respectively, we have (see Figure 3A)

$$\begin{aligned} t_C &\sim \text{Normal}\left(a \log T_C, \sigma^2\right), \\ t_I &\sim \text{Normal}\left(a \log T_I, \sigma^2\right). \end{aligned} \tag{1}$$

The noisy evidence available to a subject on a single trial consists of random draws t_C and t_I. The ideal observer in signal detection theory decides by computing the log posterior ratio of the two alternatives based on the available evidence:

$$d = \log \frac{p\left(\text{different} \mid t_C, t_I\right)}{p\left(\text{same} \mid t_C, t_I\right)}.$$

When the log posterior ratio is positive, the observer reports "different". The log posterior ratio is equal to the sum of the log likelihood ratio and the log prior ratio:

$$d = \log \frac{p\left(t_C, t_I \mid \text{different}\right)}{p\left(t_C, t_I \mid \text{same}\right)} + \log \frac{p\left(\text{different}\right)}{p\left(\text{same}\right)}.$$

Evaluating the log likelihood ratio requires a few standard steps (and assumptions about uniform priors), but leads to the very intuitive decision rule of reporting "different" when

$$t_I - t_C > b, \tag{2}$$

where *b* is a constant that depends on the log prior ratio. When the prior probabilities are equal, *p*(different)=*p*(same), then *b*=0 and the observer is unbiased. When *b* is positive, the observer has a bias for reporting "same." Using the probability distributions in Eq. , we can compute the probability that Eq. is satisfied when the true elapsed times are T_I and T_C. The difference t_I–t_C is normally distributed with mean $a(\log T_I - \log T_C)$ and variance $2\sigma^2$. It follows the probability that Eq. is satisfied (and the observer is correct) is

$$PC_{\text{model}} = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\frac{a\left(\log T_I - \log T_C\right) - b}{2\sigma},$$

where erf is the error function. The smaller the difference between T_I and T_C, the smaller the argument of the error function (which represents a signal-to-noise ratio), and the lower probability correct.

Observers might guess randomly on some proportion of trials. If the guessing rate is g, then the probability correct predicted by the model is

$$\begin{aligned} PC_{\text{model}} &= g \cdot \frac{1}{2} + (1-g)\left(\frac{1}{2} + \frac{1}{2}\operatorname{erf}\frac{a\left(\log T_I - \log T_C\right) - b}{2\sigma}\right) \\ &= \frac{1}{2} + \frac{1}{2}(1-g)\operatorname{erf}\frac{a\left(\log T_I - \log T_C\right) - b}{2\sigma}. \end{aligned}$$

In the argument of the error function, a, k, and σ can be scaled by a common factor to produce the same outcome. In order to fit the model parameters unambiguously, we therefore rewrite this part in terms of ratios:

$$\boxed{PC_{\text{model}} = \frac{1}{2} + \frac{1}{2}(1-g)\operatorname{erf}\frac{\log T_I - \log T_C - \tilde{b}}{2\tilde{\sigma}},} \tag{3}$$

where $\tilde{b} = \frac{b}{a}$ and $\tilde{\sigma} = \frac{\sigma}{a}$. Thus, we have a three-parameter model, with the parameters being guessing rate g, response bias $\tilde{b}$, and noise parameter $\tilde{\sigma}$. In the main text, we have renamed $\tilde{b}$ and $\tilde{\sigma}$ to b and σ, respectively.

Since $\log T_I - \log T_C$ can also be written as $-\log (T_C/T_I)$, Eq. (3) expresses that for a given subject, proportion correct only depends on the ratio of the time to the current sample to the time to the interfering sample.

Model Fitting

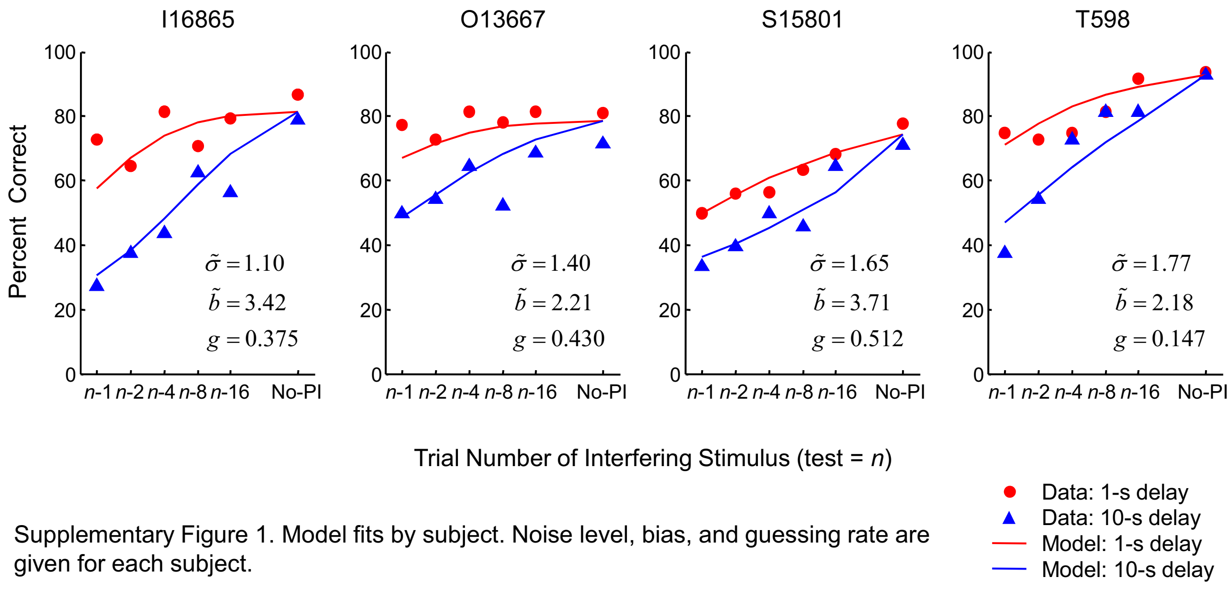
We fitted the three parameters of the model on an individual-subject basis using maximum-likelihood estimation. Experimental conditions consisted of the two delay periods (1 s and 10 s), and the six possible values of m, the number of trials the interfering sample was presented before the test (1, 2, 4, 8, 16, and ∞ for the no-PI condition). Although T_C is the same for each subject (1 s or 10 s), T_I differs between subjects due to variations in the time spent on pecking the sample display, and to a lesser extent, in response times. We used median times. The likelihood of the model parameters is the probability of the data given those parameters. The data are given by the proportion of trials on which the subject reported "same", versus "different", in each experimental condition, labeled by an index *i*. The log likelihood of the parameters is then

$$L\left(g,\tilde{b},\tilde{\sigma}\right)=\log p\left(\text{data}\,|\,g,\tilde{b},\tilde{\sigma}\right)$$
$$=\sum_{\text{conditions } i} PC_i \log PC_{\text{model, } i}\left(g,\tilde{b},\tilde{\sigma}\right)+\sum_{\text{conditions } i}\left(1-PC_i\right)\log\left(1-PC_{\text{model, } i}\left(g,\tilde{b},\tilde{\sigma}\right)\right).$$

where the model prediction $PC_{\text{model},i}$ is obtained from Eq. (3) with the values of T_I and T_C in the i^{th} experimental condition. We maximized the log likelihood of the parameters using fminsearch in Matlab.

The average maximum-likelihood value of the guessing rate was g=0.37 ± 0.08, of the familiarity bias $\tilde{b}=2.88$ ± 0.40 (indicating that pigeons had a prior favoring "same"), and of the noise parameter $\tilde{\sigma}=1.48$ ± 0.15. The model fit the individual subjects' results very well (Fig. S1) as well as the mean of the individual subject fits (Figure 3B).

Figure S1



Supplementary Figure 1. Model fits by subject. Noise level, bias, and guessing rate are given for each subject.