# A pupillary index of susceptibility to decision biases

Eran Eldar[a,b], Valkyrie Felso[c,d], Jonathan D. Cohen[c,d], Yael Niv[c,d]

[a] Psychology and Cognitive Sciences Departments, The Hebrew University of Jerusalem
[b] Max Planck University College London Centre for Computational Psychiatry and Ageing Research
[c] Princeton Neuroscience Institute, Princeton University
[d] Psychology Department, Princeton University

# Supplementary Information

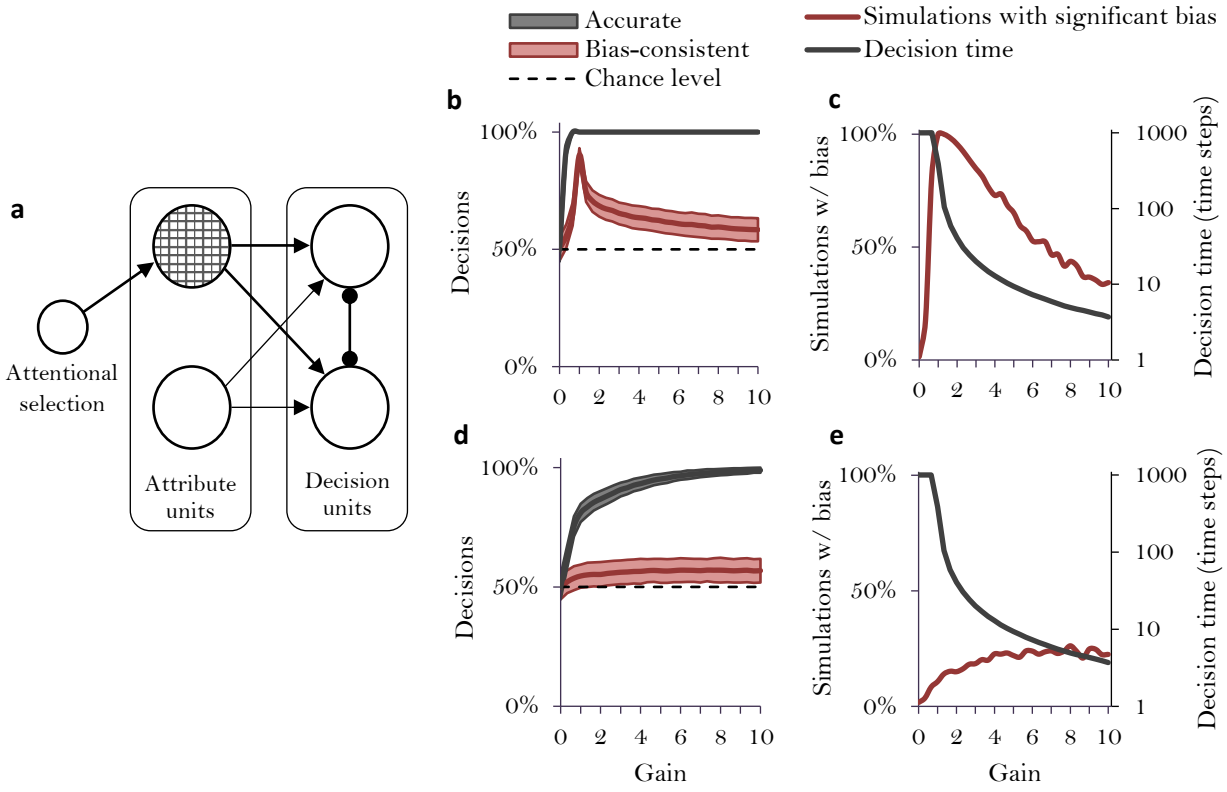
**Contents**

**Supplementary Note 1.** We have shown that high pupillary responses are associated with behavioural and fMRI findings, including broader integration in a variety of information processing tasks, that are predicted by low levels of neural gain – that is, by a reduction in the impact of all inputs on post-synaptic neural responses[1–5]. Thus, to demonstrate how large pupil responses may be related to the formation of decision biases, we simulate the effect of low gain on a framing bias in a multi-alternative, multi-attribute decision problem, using a previously developed model of such decision problems (**Supplementary Figure 1a**)[6].

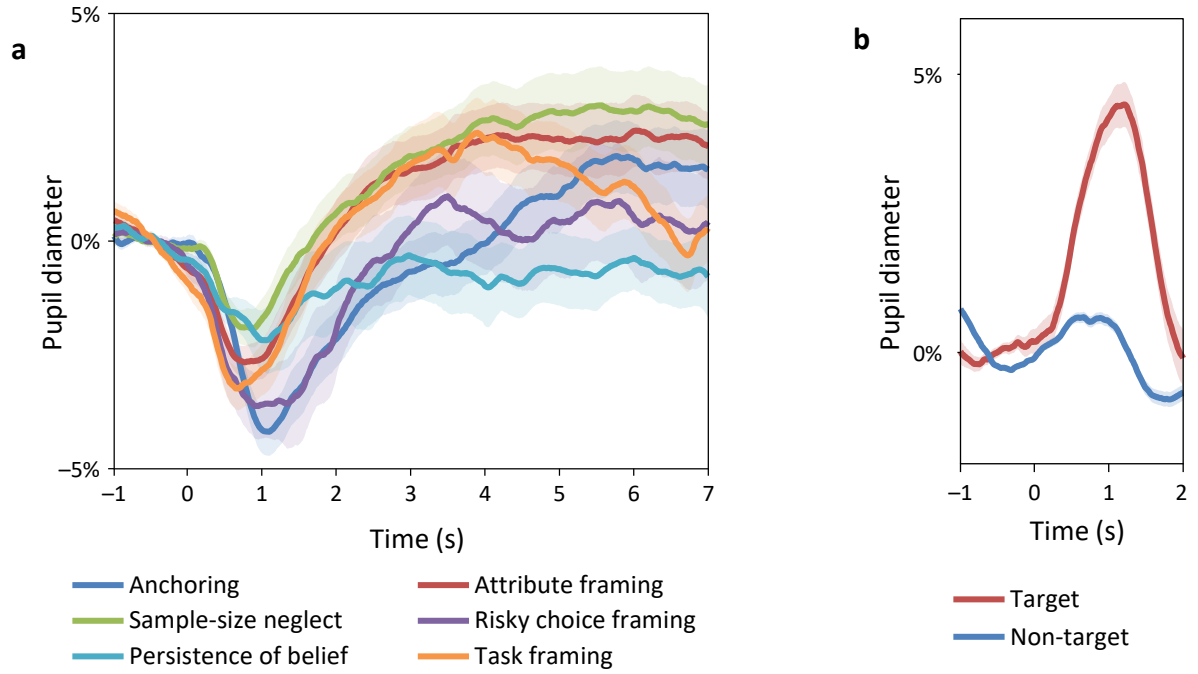
Consider a choice between two vacation destinations: Destination A has gorgeous beaches and coral reefs but a high petty crime rate, whereas Destination B has average beaches and an average crime rate. The model assumes that on each time step, attention selects one of the attributes (i.e., 'beaches' or 'crime rate') at random, and the evidence this attribute provides is accumulated at the decision layer. Since framing effects are typically conceptualized as attentional biases that exert their effect during the decision process (Levin et al., 1998), they can be naturally implemented in the model as a tendency to select certain attributes more often than others. For instance, framing the question as to where *to go* would direct attention to more frequently select the positive attribute (beaches), whereas framing the question as where *not to go* would draw attention towards the negative attribute (crime rate).

If decision making unfolds over many time steps, even a small bias can accumulate and determine the result of the decision process. In contrast, if the decision is made after only a small number of time steps, the effect of the bias would be minimized. This is where neural gain comes in: lower gain diminishes the effect of each piece of evidence on the decision process, increasing the number of time steps required to reach a decision. As a result, the effect of biases is stronger. In this way, differences in neural gain may underlie an association of framing effects with high pupil dilation (indicative of low gain) and longer decision times (**Supplementary Figure 1b,c**). We note that similar results would be obtained by any decision model that involves gradual integration of information, as long as the signal-to-noise ratio is high enough to allow reliably accurate decisions on trivial decision problems. In more noisy settings, weak biases are difficult to detect with any level of gain (**Supplementary Figure 1d,e**).

More generally, we note that the enhancement of weak influences by low gain illustrated in this model is not dependent on an evidence accumulation process being involved, as we have also previously illustrated it in decision models that do not involve evidence accumulation[3,4]




**Supplementary Figure 1.** An illustration of the effect of gain on the manifestation of a decision bias in a previously published evidence integration model of multi-attribute decisions[6]. Based on previous work[6–9], a framing bias is implemented in the model as a tendency to attend to one attribute more frequently than the other. (**a**) The model consists of two competing accumulators, one for each item. Every time step, activity $a_i$ of accumulator $i$ is updated to reflect evidence in favor of the respective item by: $\Delta a_i = 0.05\left(-a_i + g(I_i - |a_j|^+) + \epsilon\right)$, where $g$ reflects the level of gain, $I_i$ is the evidence-based excitatory input to accumulator $i$, $j$ is the competing accumulator whose positive component ($|\,|^+$) provides inhibitory input, and $\epsilon$ is zero-mean normally-distributed noise with a standard deviation of 0.5. On each time step, one attribute is randomly selected, and the evidence in favor of each item is accumulated by the competing decision units. One of the attributes favors one item, and thus generates input of 1.2 to one accumulator and 0.8 to the other accumulator. The other attribute favors the other item, and thus generates the reversed input. Bias was implemented as a tendency to select one of the attributes more frequently ($p = 0.55$). A decision is reached once one of the accumulators reaches a value of 1 or 1000 time steps are completed. (**b**) Accuracy and bias as a function of gain. Average proportion (± standard deviation) of accurate and bias-consistent decisions in a sample of 100 simulated decisions. Accuracy was measured in a separate set of simulations in which both attributes favored the same choice. The decision process was simulated 100,000 times with each level of gain. The dashed line indicates chance-level accuracy and bias. With low (but not too low) gain accuracy is high, but so is the effect of the bias. (**c**) Bias detectability and mean decision time as a function of gain. Bias detectability was measured as the proportion of simulations in which a statistically significant bias ($p<0.05$, binomial test) was evident, with each simulation including 100 decisions. Bias is associated with *slower* rather than faster decisions, due to low gain. (**d**, **e**) Same as panels **b** and **c**, but with a high level of noise (standard deviation = 2), set such that accuracy would be less than optimal. In this case, biases are difficult to detect with any level of gain.




**Supplementary Figure 2.** Pupillary responses during each of the decision-making tasks (**a**; $n = 44$) and the auditory oddball task (**b**; $n = 6$). Responses were computed based on the same number of trials per subject to be included in the experiment. Time 0 denotes trial onset. Measurements were normalized by subtracting the average 1s pre-trial baseline diameter, and dividing by each individual's pre-experiment reference measurement. Shaded area: S.E.M.

**Supplementary References**


1. Servan-Schreiber, D., Printz, H. & Cohen, J. D. A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science* **249**, 892-895 (1990).
2. Aston-Jones G. & Cohen J. D. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* **28**, 403-450 (2005).
3. Eldar, E., Cohen, J. D. & Niv, Y. The effects of neural gain on attention and learning. *Nat. Neurosci.* **16**, 1146-1153 (2013).
4. Eldar, E., Niv, Y. & Cohen, J. D. Do you see the forest or the tree? Neural gain and breadth versus focus in perceptual processing. *Psychol. Sci.* **27**, 1632–1643 (2016).
5. Eldar, E., Cohen, J. D., & Niv, Y.. Amplified selectivity in cognitive processing implements the neural gain model of norepinephrine function. *Behav. Brain Sci.* **39**, e206 (2016).
6. Usher, M. & McClelland, J. L. Loss aversion and inhibition in dynamical models of multialternative choice. *Psychol. Rev.* **111**, 757 (2004).
7. Levin, I. P., Johnson, R. D., Russo, C. P. & Deldin, P. J. Framing effects in judgment tasks with varying amounts of information. *Organ. Behav. Hum. Decis. Process.* 36, 362-377 (1985).
8. Keren, G. On the definition and possible underpinnings of framing effects: A brief review and a critical evaluation. In G. Keren (Ed.), *Perspectives on framing* (pp. 3–33). New York: Psychology Press (2011).
9. Kreiner, H. & Gamliel, E. The Role of Attention in Attribute Framing. *J. Behav. Decis. Mak.* DOI: 10.1002/bdm.2067 (2017).