**Supplementary Material**

Attention Induces Conservative Subjective Biases in Visual Perception

Dobromir Rahnev, Brian Maniscalco, Tashina Graves, Elliott Huang, Floris P. de Lange, Hakwan Lau


Supplementary Figures and Methods:

1. Supplementary Figure 1. Attentional cuing leads to conservative detection biases

2. Supplementary Figure 2. Results from a control experiment that involved eye tracking

3. Supplementary Figure 3. The influence of the unprobed stimulus on criterion in Experiment 1

4. Supplementary Figure 4. Hit and false alarm rates in Experiment 2

5. Supplementary Figure 5. A depiction of the signal detection theoretic model that accounted for our findings

6. Supplementary Figure 6. Model fits

7. Supplementary Figure 7. Models in which attention does not affect the variability of the distributions cannot account for the observed data

8. Supplementary Figure 8. Distributed attention leads to disproportionately high visibility ratings

9. Supplementary Methods
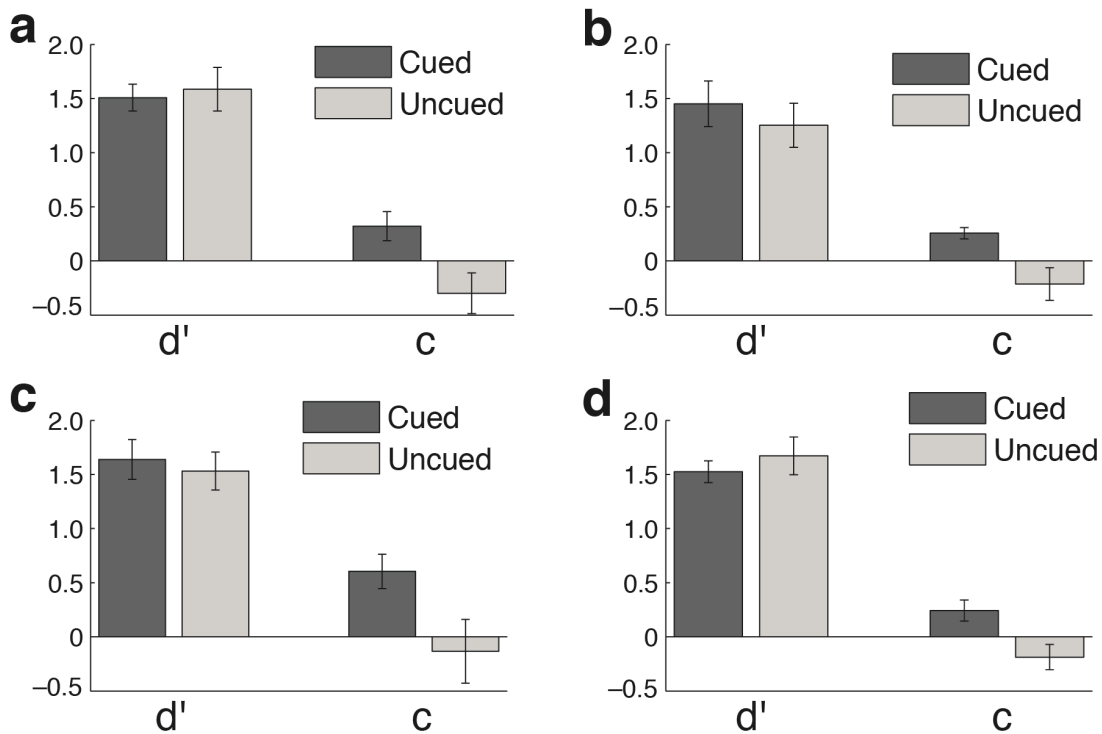
Supplementary Figure 1. Attentional cuing leads to conservative detection biases. To investigate the robustness of the results in Experiment 1, we replicated the experiment four additional times while changing different aspects of the design. As in Experiment 1, in all four experiments we titrated the contrast of the gratings so that subjects performed equally well in the cued and uncued trials (had the same *d'*, all *p*'s > 0.2). In all four additional experiments subjects were still more conservative in detecting the cued gratings (as indicated by a higher criterion *c*). The difference between the criterion for detecting cued and uncued gratings was significant (**a**) when we encouraged subjects to respond optimally by giving them additional monetary reward, as well as an explicit payoff structure that promoted unbiased responding ($t(5) = 3.10$, $p = 0.027$), (**b**) when in addition to the monetary reward and the payoff structure, subject were given trial-by-trial feedback which is thought to diminish suboptimal perceptual biases ($t(9) = 3.08$, $p = 0.013$), (**c**) when the stimulus presentation was decreased to 50 ms to avoid eye movements ($t(4) = 3.65$, $p = 0.022$), and (**d**) when eye movements were monitored to ensure that subjects were indeed fixating as instructed ($t(5) = 3.45$, $p = 0.018$). This and the subsequent figures show means ± s.e.m.

2
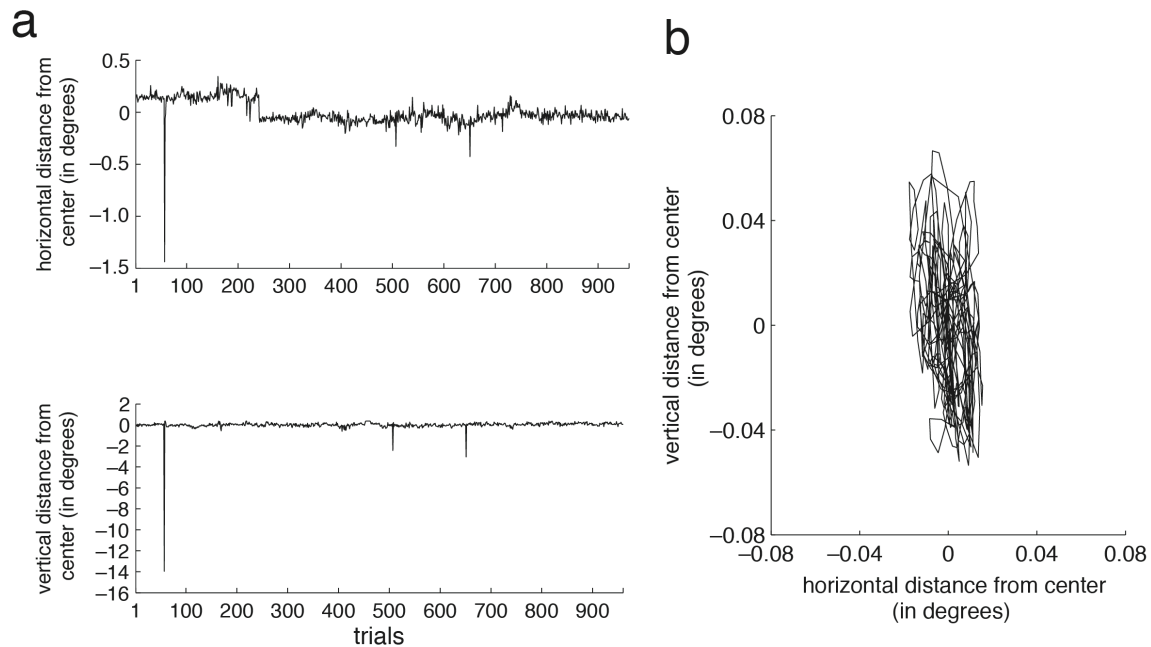
Supplementary Figure 2. Results from a control experiment that involved eye tracking. (**a**) We plotted the average horizontal and vertical eye position during the stimulus presentation for each of the 960 trials for one representative subject who showed variability closest to the group average (i.e., not the "best" subject). The standard deviation of the fixation positions for this subject was 0.15 degrees for the horizontal dimension and 0.67 for the vertical dimension. The average standard deviations across the 6 subjects were 0.23 and 0.55 for the horizontal and vertical dimensions, respectively. We also found that subjects made eye-movements larger than 1 degree on only 1.2% of the trials (stimuli were located 5 degrees away from fixation). To ensure that eye-movements did not contribute to our results on criterion, we correlated across subjects the propensity to make eye movements with the magnitude of the criterion difference between cued and uncued trials. It turned out that the correlation was not significant and it was, if anything, in the opposite direction such that less eye movement was associated with bigger criterion difference ($r = -0.36$, $p = 0.49$). Thus, eye movements could not account for the effects of attention on criterion. (**b**) We show the fixations for one trial from the subject from (a).
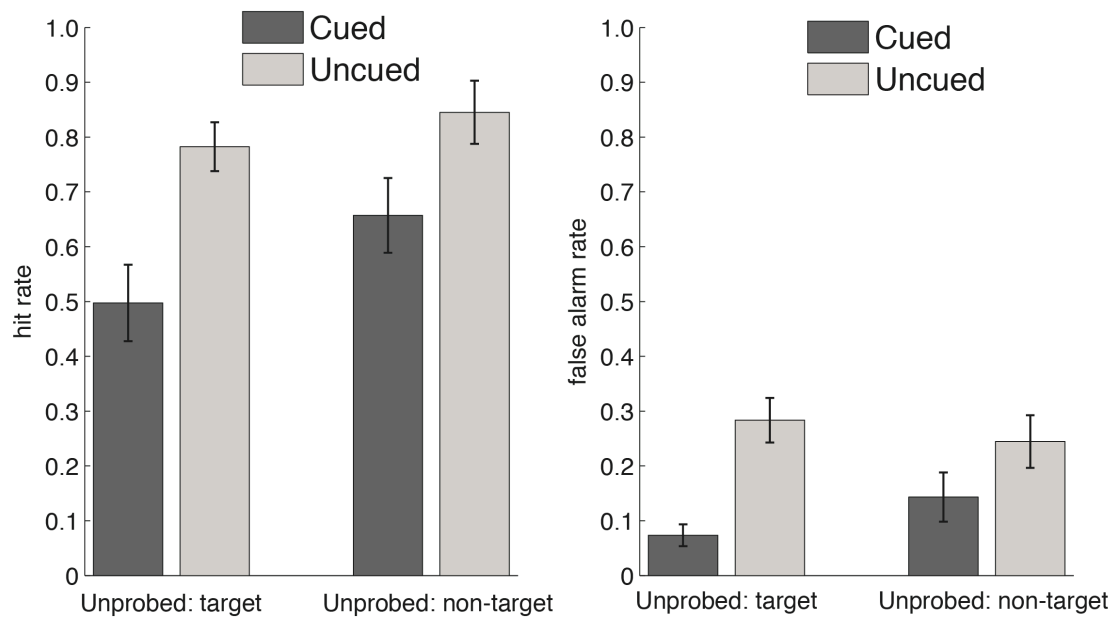
3

Supplementary Figure 3. The influence of the unprobed stimulus on criterion in Experiment 1. To investigate the putative role of the stimulus in the "unprobed" location (i.e., the stimulus that was not response-cued), we plotted hit and false alarm rates separately for the cases when the unprobed location contained a target or a non-target. The results showed that hit and false alarm rates were higher in the uncued condition regardless of the identity of the unprobed location. We also found an interaction between attentional cuing and the identity of the unprobed stimulus: when a noise patch was present in the unprobed location the difference between the cued and uncued location decreased in terms of both hit rate ($F(1,8) = 10.81$, $p = 0.011$) and false alarm rate ($F(1,8) = 7.78$, $p = 0.024$). Therefore, it appears that the identity of the unprobed location influenced subjects' detection bias. Nevertheless, this influence could not account for the overall decrease of the criterion c with attention.

4
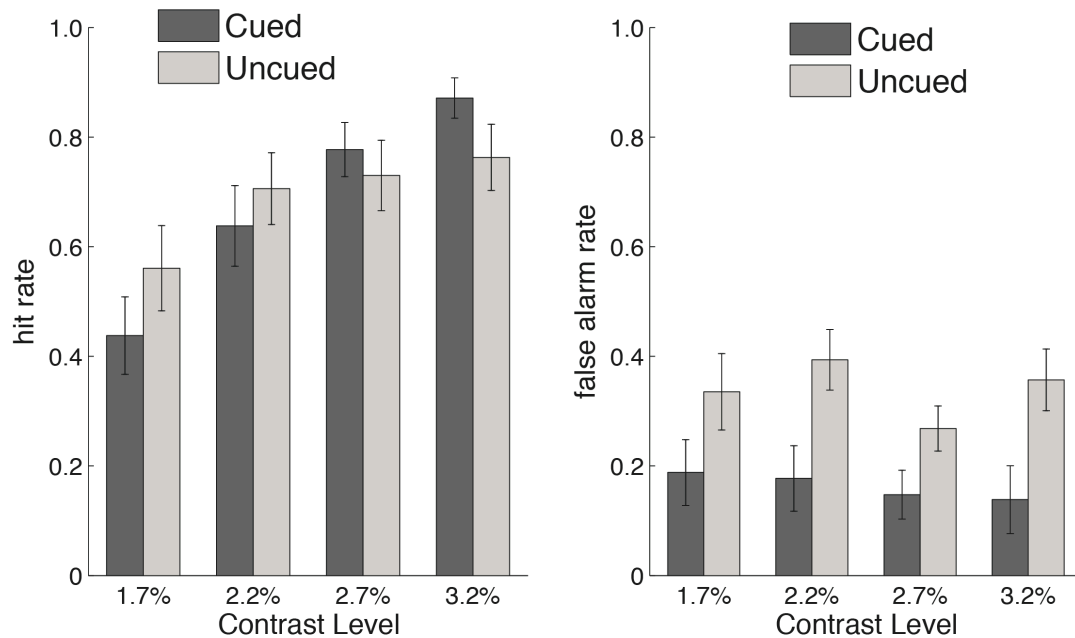
Supplementary Figure 4. Hit and false alarm rates in Experiment 2. We plotted the hit and false alarm rates for each contrast level for the cued and uncued trials separately. A multiple regression with factors of attention, contrast, and subject-specific effects demonstrated that contrast level modulated hit rates ($p < 0.001$) but not false alarm rates ($p = 0.48$). These effects are in agreement with our signal detection theoretic model (see **Supplementary Fig. 5**): the model postulates that subjects use the same unified criterion for detection in both the cued and uncued locations, and that the unified criterion is fixed with respect to the mean of the Target-Absent distribution. Thus, according to our model, false alarm rates should stay approximately constant across contrast levels for both cued and uncued stimuli (the uncued stimuli should have a higher false alarm rate because of the extra noise in the distributions).
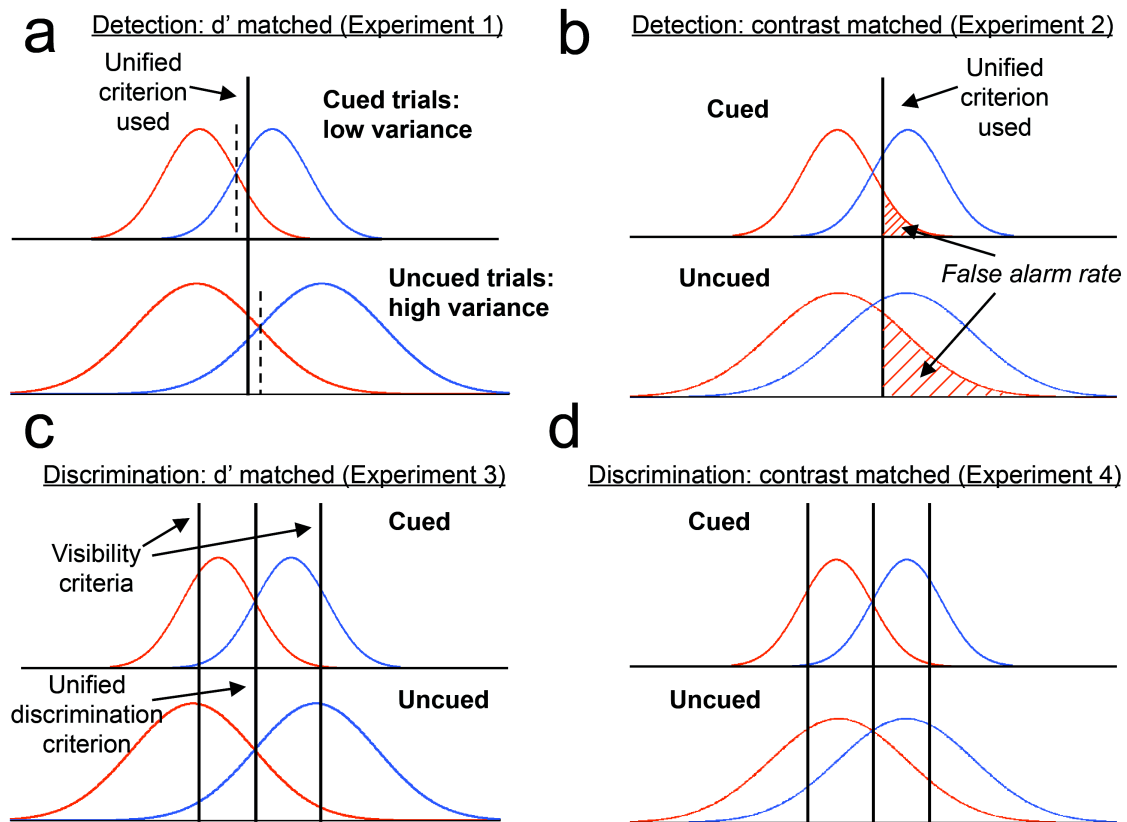
Supplementary Figure 5. A depiction of the signal detection theoretic model that accounted for our findings. The model postulates that attention leads to smaller trial-by-trial variability in the internal perceptual signal. Here we try to give an intuitive account of how the model is able to explain the results of the four experiments. (**a**) In Experiment 1 we equated the detection sensitivity (*d'*) of the cued and uncued stimuli by presenting higher contrast stimuli in the uncued locations. We found that subjects produced higher hit and false alarm rates for the uncued stimuli (see **Fig. 2**). In our model the higher contrast of the uncued stimuli was represented by a bigger distance between the Target-Absent (red; left) and Target-Present (blue; right) distributions. This larger difference between the cued and uncued trials was counterbalanced by the larger variability of the distributions for the uncued trials, so that *d'* (signal-to-noise ratio) for cued and uncued trials was equated. Our model also assumes a single unified criterion used for both cued and uncued trials (previous studies have shown that subjects indeed tend to use a single unified criterion for detection of targets in two possible locations, compromising optimality for both[1]). As depicted in the figure, these assumptions lead to conservative detection for the cued stimuli and liberal detection for the uncued stimuli, which is what the experimental data showed. The optimal (unbiased) criteria for both conditions are depicted with dashed lines. (**b**) In Experiment 2, contrast was kept the same for cued and uncued stimuli (resulting in a higher *d'* value for the cued stimuli). Subjects were more conservative in detecting the cued stimuli, with the effect being largest for low stimulus contrasts. In our model, unlike (a) where the distance between the peaks of the distributions was larger for uncued stimuli (because of the difference in contrast between

6

cued and uncued stimuli), in Experiment 2 one would expect this distance to be higher for cued stimuli because attention leads to boosting of the perceptual signal. Our results showed that subjects tended to avoid excessive false alarms by fixing their unified detection criterion with respect to the Target-Absent distribution to achieve a relatively constant level of false alarms[2] (see also **Supplementary Fig. 4**). Because the Target-Absent distribution for uncued trials is wider, this leads to more false alarms for uncued trials (see shaded regions). However, the hit rate depends on the contrast presented (**Supplementary Fig. 4**). Specifically, for low contrast levels the Target-Present distributions for both cued and uncued stimuli are relatively close to the Target-Absent distributions and the unified criterion falls to the left of the peak of the Target-Present distribution. In this case, due to the higher variability of the Target-Present distribution for uncued stimuli, hit rate would be higher for uncued stimuli. However, for higher contrast levels the Target-Present distribution's peak would be to the right of the unified detection criterion which would lead to lower hit rate for uncued stimuli. In the figure the unified criterion falls on the left of the peak of the Target-Present distributions which leads to higher hit rate for cued stimuli (i.e., a condition similar to using high contrast). (**c**) In Experiment 3, we again equated the discrimination sensitivity d' and observed higher visibility ratings for uncued stimuli. According to signal detection theory, a discrimination task can be represented with Gaussian distributions as in (a) and (b). In this case, the Target-Present and Target-Absent distributions would correspond to the distributions for right- and left-tilted patches. To give visibility (or confidence) ratings in a discrimination task, subjects can set additional "visibility" criteria next to the unified criterion for discrimination. High visibility ratings will be given for any trial that produces an internal response that falls to the right of the rightmost visibility criterion or to the left of the leftmost visibility criterion. If the relative lack of attention leads to larger variability of internal perceptual signal for the uncued stimuli, the wider spread means that more trials will exceed the visibility criteria for high visibility ratings. Since the two visibility criteria behave similarly, in the following discussion we focus on the rightmost visibility criterion. For that criterion, high confidence trials coming from the Target-Present distribution correspond to hits in the detection task depicted above, while high confidence trials coming from the Target-Absent distribution correspond to false alarms in the detection task. Therefore, in the model the uncued stimuli would produce higher visibility ratings for the same reasons for which uncued stimuli receive more hit and false alarms in (a). (**d**) Finally, in Experiment 4 we used the same contrasts for the cued and uncued stimuli. As in Experiment 2, this resulted in a higher d' value for the cued stimuli. Also, similarly to (b), in this experiment one would expect the distance between the two Gaussian distributions to be higher for cued stimuli because attention boosts the perceptual signal. Finally, according to our model, as in (b) subjects would give more high visibility ratings (graphically corresponding to hit and false alarms in the detection experiment in (b)) for the uncued stimuli when the contrast is low but not when the contrast is high, which is again what we found experimentally.
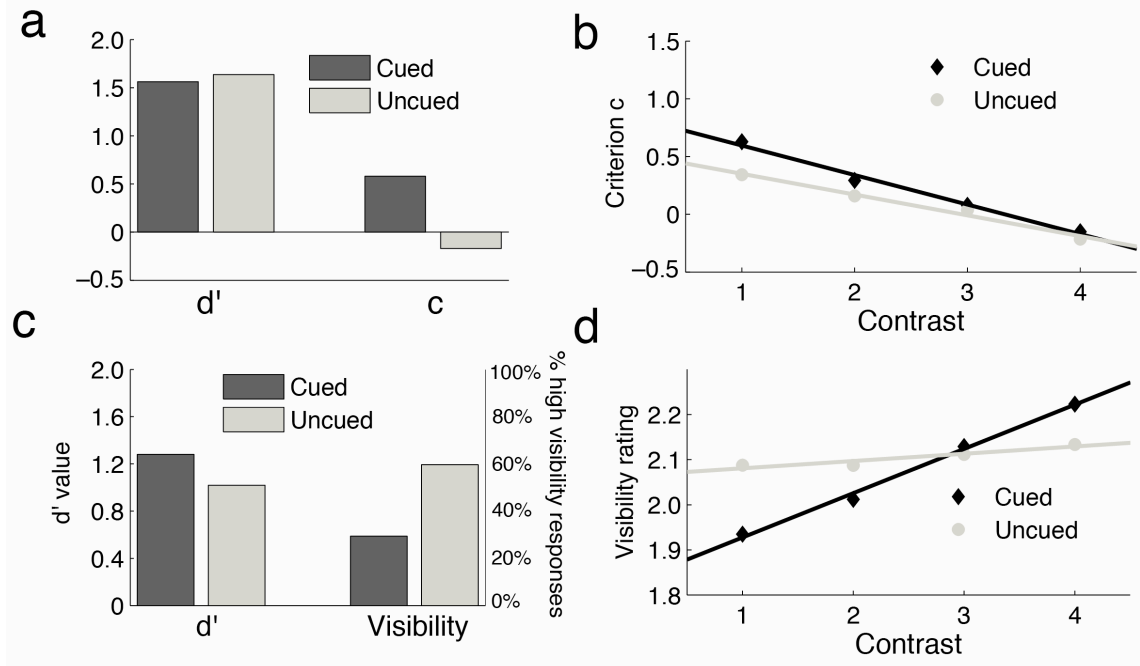
Supplementary Figure 6. Model fits. We constructed a signal detection theoretic model in which attention increased the magnitude of the perceptual signal and decreased its trial-by-trial variability (see **Supplementary Methods** for details). We fit the model to the data for each observer and plotted the mean predicted value across observers for each of Experiments 1–4. The model was able to capture the pattern of the data in all experiments (see **Fig. 2** for comparison). In (**b**) and (**d**) we marked the contrasts with the numbers 1 to 4 where 1 stands for the lowest presented contrast and 4 stands for the highest contrast. Based on the model fits across the four experiments, attention seemed to reduce the variability (standard deviation) of the cued distributions by a factor of ~2. We tested our intuition that the critical feature of the model was indeed this reduction in noise (see **Supplementary Fig. 5**). We investigated if models in which attention does not affect the variability of the signal were able to explain the data. The results showed that the model in which attention modulated both the mean and SD of the distributions ('mean+SD' model) outperformed models in which attention only modulated the mean of the distributions ('mean-only' model) or had no effect ('null' model). This was true even when we used model selection methods (Bayesian information criterion; BIC) to punish models for complexity. Thus our model provided good fits not just because it has more parameters. The averaged posterior probabilities (estimated from BIC) for the 'mean+SD' model were 0.91, 0.77, 0.56, and 0.93 for Experiments 1–4, respectively. These values were higher than the values for the 'mean-only' (0.04, 0.06, 0.20, and 0.07) or the 'null' (0.05, 0.17, 0.24, and 0.00) models, providing strong evidence that the attentional modulation of the signal variability is indeed the critical feature of our model.
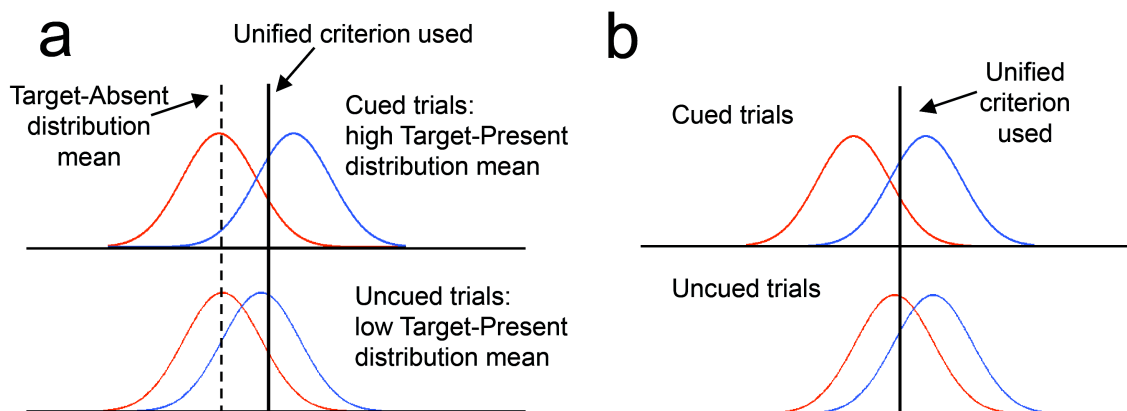
8

Supplementary Figure 7. Models in which attention does not affect the variability of the distributions cannot account for the observed data. As discussed in the legend of **Supplementary Figure 6**, the 'mean-only' model did not provide good model fits. Here we give an intuition as to why that model could not fit the observed data. We focus on the detection experiments, though similar argument could be given for the discrimination experiments (**a**) A depiction of the 'mean-only' model; the Target-Absent distribution is shown in red (left), while the Target-Present distribution is shown in blue (right). The 'mean-only' model allowed for attention to increase the mean of the Target-Present distribution, without changing the mean of the Target-Absent distribution. The assumption that attention does not alter the mean of the Target-Absent distribution is based on the underlying neuronal physiology. Indeed, when a non-target is presented in the uncued condition, we expect the firing to be near baseline so that attention would not have much room to further decrease the firing. Given the above assumptions, the 'mean-only' model predicts that cued and uncued trials will have similar levels of false alarms, and that cued trials should have a higher hit rate than uncued trials. However, Experiments 1 and 2 showed a very different pattern of results. Specifically, the false alarm rate was higher for uncued trials (see **Supplementary Figs. 3** and **4**). Hit rates were also higher for the uncued trials in both Experiment 1 and the first two contrast levels in Experiment 2, which is the opposite of what the 'mean-only' model would predict. (**b**) A depiction of an alternative version of the 'mean-only' model which can mathematically account for the observed results. This alternative model allows attention to move both the Target-Absent and Target-Present distributions in either direction. In order for this model to account for the findings, in Experiment 1 attention would need to shift the Target-Absent distribution 0.7 standard deviations to the left. (It is impossible to analyze the influence of attention on the Target-Present distribution because the contrast differed between the cued and uncued stimuli.) In Experiment 2, attention would need to shift the Target-Absent distribution 0.6 standard deviations to the left. The Target-Present distribution would need to move 0.4 standard deviations to the left for the lowest two contrasts, and 0.03 and 0.4 to the right for the highest two contrasts, respectively. Thus, in order for the alternative 'mean-only' model to fit the data, the overall effect of attention would need to be to decrease the means of the Target-Absent and Target-Present distributions. This effect would correspond to attention having a largely suppressive influence on neuronal firing. However, such an effect does not fit with the well documented findings that attention boosts the firing rate. Even for the highest

9

contrast in Experiment 2, the boost in neuronal firing for the Target-Present distribution would be smaller than the suppression for the Target-Absent distribution. Further, this kind of alternative 'mean-only' model would not be able to account for our results from the discrimination studies, since a move of both the distributions to the left (or to the right) would result in a bias for choosing either left- or right-tilted gratings and such a bias was not observed in the data. Therefore, because of its neural implausibility and lack of generalizability to all our data, we did not implement this alternative 'mean-only' model computationally.
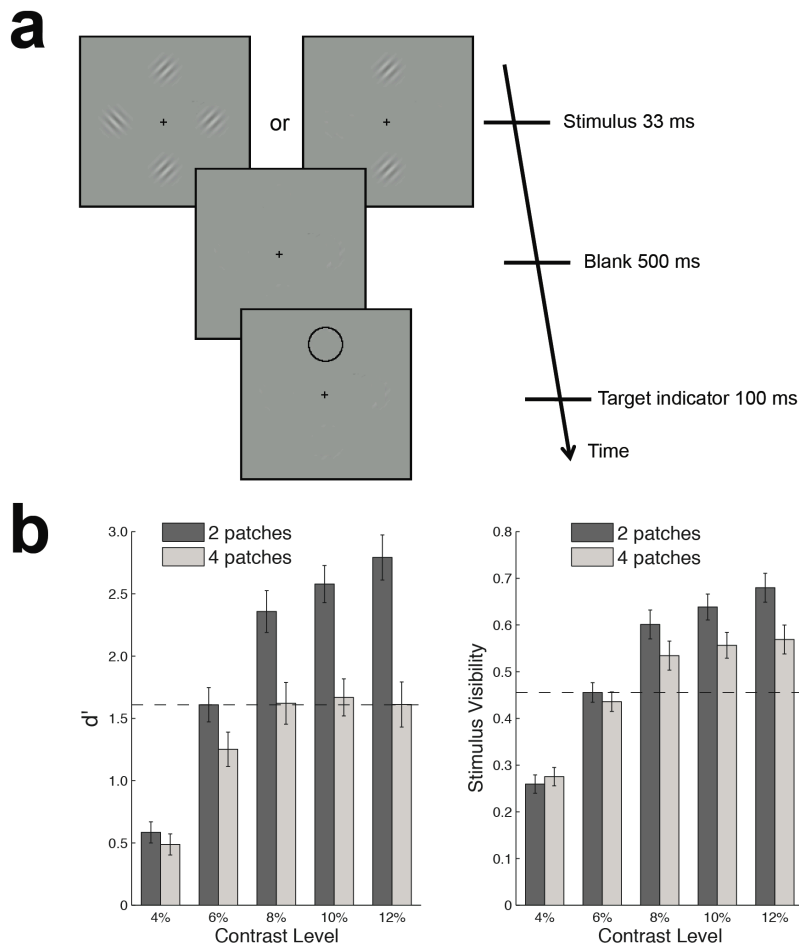
10

Supplementary Figure 8. Distributed attention leads to disproportionately high visibility ratings. (**a**)We presented different number of stimuli on the screen in order to manipulate how subjects distribute their attention to different objects. In one condition we used 2 items on the screen (a relatively focused mode of attention), and in the other we used 4 items on the screen (a relatively distributed mode of attention). The stimuli were presented for 33 ms. After a delay of 500 ms, subjects saw a response cue that instructed them on which stimulus they should do the task. Subjects had to indicate the tilt (left/right) of a Gabor patch and rate the visibility (high/low) of the tilt of that patch. Subjects completed 8 blocks of 125 trials each for a total of 1000 trials. Within each block there were always either 2 or 4 patches. We computed d' and stimulus visibility for each of the 5 levels of contrast that we used in this experiment. (**b**) Both d' and visibility ratings increased with higher contrast. Further, d' was roughly similar for the 2-patch task with 6% contrast, and the 4-patch task with 8% ($p = 0.95$), 10% ($p = 0.7$), and 12% ($p = 0.99$) contrast (see the horizontal dashed line). Nevertheless, compared to the 6%-contrast 2-patch task, the visibility of the grating was judged to be higher for the 4-patch task for the 8% contrast ($p = 0.01$), 10% contrast ($p = 0.01$), and 12% contrast ($p = 0.02$). Thus, similar to the matched-d' discrimination experiment in the main text (Experiment 3), less attention (in the 4-patch task) led to higher subjective stimulus visibility ratings even though discrimination sensitivity (d') was matched. Lower values of stimulus visibility indicate a less visible stimulus.

11

Supplementary Methods

Participants
Fifty-eight subjects in total (30 women; mean age = 23; range = 18–35 years) participated in eight psychophysical experiments (Experiments 1–4 and four additional experiments that served as control studies for Experiment 1; we call these experiments Control Experiments 1–4). Three subjects participated in two of the experiments, two participated in three of them, and one participated in four. All subjects were naive regarding the purposes of the experiments, had normal or corrected-to-normal vision, and signed an informed-consent statement approved by the local ethics committee

Materials and Procedure
In the detection experiments (Experiments 1 & 2), stimuli were presented on a gray background (6.0 cd/m^2). Four circles (5° visual angle in diameter) were presented in the four quadrants of the screen with the center of each circle located 5° away from fixation. This configuration was chosen in order to minimize eye movements, as in previous studies[3]. A white (27 cd/m^2) arrow pointed to one of the two diagonals. Subjects were seated in a dimmed room about 60 cm away from the computer monitor and instructed to maintain fixation at a central red dot (0.4°) displayed on top of the white arrow for the duration of each trial. Stimuli were generated using Psychophysics Toolbox[4] in MATLAB (MathWorks, Natick, MA) and were shown on an iMac monitor (19 inch monitor size, 1680 x 1050 pixel resolution, 60 Hz refresh rate).

The stimuli consisted of a noisy background composed of uniformly distributed intensity values (8% contrast). On top of the noise, we added gratings (0.5 cycles/degree) with probability of 50%. The appearance of the gratings was independent for the cued and uncued diagonals: thus gratings could appear in both diagonals, in neither, or in just one diagonal.

The trials in the main task of the experiment began with 500 ms of a centrally presented pre-cue in the form of an arrow. This pre-cue indicated the likely location of target appearance. After 500 ms, four circles were presented for 367 ms (**Fig. 1**; except for Control Experiment 3 where the stimuli were presented for 50 ms). The pre-cued diagonal alternated in blocks of 40 trials. This "blocking" of spatial attention reduced the cognitive demands on the subjects and made it less likely that they could confuse the identity of the cued diagonal. Subjects were asked to indicate whether a grating was present in the diagonal that was highlighted with a response cue. The response-cued diagonal was the same as the pre-cued diagonal on 70% of the trials. Subjects were informed about this fact, which encouraged them to use more attentional resources for the processing of the stimuli in the cued compared to the uncued diagonal.

In Experiment 1 and Control Experiments 1–4 (which served as controls to Experiment 1) the contrast of the stimuli was adjusted online to achieve equal d' for the cued and uncued trials. In Experiment 2, the contrast for each subject was fixed throughout the experiment. We used the QUEST staircase procedure[5] to find the contrast level for each subject that would produce about 90% correct responses for the cued and uncued stimuli altogether

(mean contrast = 2.47%, SD = 0.9%). Each of the 4 runs in this experiment included gratings of fixed contrast – 50%, 67%, 83%, and 100% of the originally obtained contrast value, respectively. The order of the runs was randomized between participants.

Control Experiments 1 and 2 included an explicit pay-off structure that encouraged unbiased responding. Subjects were given 1 point for each correct answer (hit or correct rejection) and 0 points for each incorrect answer (false alarm or miss). Control Experiment 2 further included trial-by-trial feedback. In order to further increase subjects' motivation to perform the detection optimally, in each of the two experiments we awarded an extra $10 to the two subjects with the highest scores. In Control Experiment 3 the stimuli were presented for 50 ms to minimize the possibility of eye movements between the stimuli. Finally, in Control Experiment 4 we measures eye movements explicitly using an EyeLink 1000 (SR Research, Osgoode, ON, Canada) infrared camera recording at 1200 Hz. The eye-tracker had gaze resolution of 0.01° (noise limited) and gaze position accuracy of 0.5°.

The discrimination experiments (Experiments 3 and 4) were similar to the detection experiments. The main difference was that subjects were asked to indicate whether the tilt of the gratings (which were always presented) was 45° or 135°, and then indicate the visibility of the tilt of the gratings. In Experiment 3 we used a 2-point scale (high/low), while in Experiment 4 we used a 4-point scale (1 – not visible at all; 4 – highly visible). We were careful in explaining to the subjects that they should rate the visibility of the tilt of the grating rather than the overall brightness of the stimulus. In Experiment 3 the contrast of the cued and uncued stimuli was updated online as in Experiments 1 and Control Experiments 1–4. In Experiment 4 we chose fixed levels of contrast but unlike Experiment 2, for simplicity the contrasts here were chosen to be the same for all participants (1.7, 2.2, 2.7, and 3.2% contrast) and were not separated in different runs.

In each of Experiments 1–4 and Control Experiments 1–4, subjects completed 960 experimental trials separated into 4 runs of 6 blocks. Feedback was given at the end of each block consisting of 40 trials, except for Control Experiment 2 where feedback was given after each trial.

Our task was relatively demanding and our subjects were untrained in psychophysical tasks. Overall seven subject needed to be excluded because of inability to perform better than chance (0, 0, 3, and 1 subject was excluded subjects from Experiments 1–4, respectively; 0, 2, 1, 0 subjects were excluded from Control Experiments 1–4, respectively).

Statistics
We used standard statistical techniques such as analysis of variance (ANOVA) and t tests. For the psychophysical experiments we computed the signal detection measures d' and criterion c by calculating the hit rate (HR) and false alarm rate (FAR). Then $d' = z(HR) - z(FAR)$ and $c = -0.5*(z(HR)+z(FAR))$, where z is the inverse of the normal distribution function[2].

13

Computational Modeling Assumptions

We modeled the behavioral results from Experiments 1–4 with a computational model based on signal detection theory (SDT). We made several standard assumptions: (1) the two stimuli used in the experiment gave rise to internal signals normally distributed along some decision axis; (2) perceptual decisions were made by comparing the signal on some decision axis to a criterion; (3) confidence judgments were made by comparing the signal on some decision axis to multiple criteria, corresponding to the multiple confidence ratings available to subjects in this experiment; and (4) criteria for perceptual decisions and confidence ratings were set in the same way for cued and uncued stimuli. The last assumption derives from previous research[1] which has demonstrated that subjects tend to use a single set of criteria for different sets of stimuli even if they are clearly labeled and spatially separated.

Model specifications

In our SDT model attention modulated both the signal and the noise of the internal representations. Thus, in this model attention changed both the distance (μ) between the Gaussian distributions and the standard deviation (σ) of the distributions (hence, we refer to this model as 'mean+SD'). The standard deviation for the uncued stimuli (σ_{uncued}) was always set to 1. When the model was applied to Experiment 1, it included four free parameters: σ_{cued}, μ_{uncued}, a parameter that quantified the increase of μ_{uncued} with attention, and the location of the detection criterion. When applied to Experiment 2, it included seven free parameters: σ_{cued}, $\mu_{uncued\_contrast}$ for each of the four levels of contrast (4 parameters), a parameter that quantified the increase of each μ_{uncued} with attention, and the location of the detection criterion. Applied to Experiment 3, the model contained six free parameters: σ_{cued}, μ_{uncued}, a parameter that quantified the increase of μ_{uncued} with attention, and the criteria levels used for discrimination and visibility judgments (3 parameters). Finally, when applied to Experiment 4, the model contained thirteen free parameters: σ_{cued}, $\mu_{uncued\_contrast}$ for each of the four levels of contrast (4 parameters), a parameter that quantified the increase of each μ_{uncued} with attention, and the location of the criteria levels used for discrimination and visibility judgments (7 parameters corresponding to the 8 possible answers; the 8 answers are produced by combining the 2 stimulus choices and 4 visibility levels).

To test if our model only produced good fits because of the number of free parameters, we compared it to two simpler SDT models. In the first one, attention was allowed to affect only the distance (μ) between the Gaussian distributions ('mean-only' model), while in the second attention was not allowed to modulate anything ('null' model). In all four applications, the 'mean-only' model had one fewer parameter than the 'mean+SD' model, which allowed both the mean and SD of the distributions to be modulated by attention: it lacked the σ_{cued} parameter. The 'null' model had two fewer parameters than the 'mean+SD' model: it lacked both the σ_{cued} and the parameter that quantified the increase of μ_{uncued}. All other parameters were identical across the 3 models. The point of this comparison was to show that the extra parameters in the 'mean+SD' were necessary and worth the extra complexity.

14

Following standard SDT methods, we assumed that the 2-dimensional stimulus image was reduced to a single scalar value representing the likelihood that this stimulus was a target or a non-target. We did not explicitly model this information reduction. Our SDT model operated at the level of overall probabilities of giving each possible response following the presentation of each kind of stimulus and did not address the trial-by-trial pattern of responses.

Model fitting
We fit the models to the data using a maximum likelihood estimation approach that has previously been used within a signal detection framework[6]. Briefly, the likelihood of a set of signal detection model parameters given the observed data can be calculated using the multinomial model. Formally,

$$L(\theta|data) \propto \prod_{i,j} Prob_\theta(Resp_i|Stim_j)^{n_{data}(Resp_i|Stim_j)}$$

where each $Resp_i$ is a behavioral response a subject may produce on a given trial, and each $Stim_j$ is a type of stimulus that may be shown on that trial.

The expression "$n_{data}(Resp_i | Stim_j)$" is a count of how many times a subject actually produced $Resp_i$ after being shown $Stim_j$.

The expression "$Prob_\theta(Resp_i | Stim_j)$" denotes the probability with which the subject produces the response $Resp_i$ after being presented with $Stim_j$, according to the signal detection model specified with parameters θ. According to SDT, in an experiment with two possible stimuli and n levels of confidence subjects have 2*n possible responses and therefore set 2*n-1 decision criteria that allow them to determine how to categorize each new trial. We denote these monotonically increasing criteria as $c_1, c_2, \ldots c_{2*n-1}$. In addition let $c_0 = -\infty$, and $c_{2*n} = \infty$. The ordering of the set of response types "Resp" follows the ordering of response types defined by setting the monotonically increasing set of criteria $c_1, \ldots c_{2*n-1}$ on a decision axis. For instance, in Experiments 1 and 2, $Resp_1 =$ "target absent" and $Resp_2 =$ "target present"; and in Experiment 3, $Resp_1 =$ "high visibility, left tilt", $Resp_2 =$ "low visibility, left tilt", $Resp_3 =$ "low visibility, right tilt", $Resp_4 =$ "high visibility, right tilt." Then, if we assume that $Stim_j$ gives rise to a Gaussian distribution with a mean μ_j and standard deviation σ_j, the expression $Prob_\theta(Resp_i | Stim_j)$ evaluates to:

$$Prob_\theta(Resp_i|Stim_j) = \int_{c_{i-1}}^{c_i} \frac{1}{\sqrt{2\pi\sigma_j^2}} \, e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}} \, \mathrm{d}x$$

Note that the models were not fit to summary statistics of performance such as percent correct or average visibility. Rather, models were fit to the full distribution of probabilities of each response type contingent on each stimulus type. Various kinds of summary statistics (e.g. *d', c,* percent correct, average visibility ratings, and so on) can be derived from this full behavioral profile of stimulus-contingent response probabilities.

We fit all models under consideration to the observed data by finding the maximum-likelihood parameter values θ. Maximum likelihood fits were found using a simulated annealing procedure[7]. Model fitting was conducted separately for each subject's data. The estimation procedure was reliable; subsequent repetitions of the model fitting procedure produced negligible variations in the parameter estimates for each model of each subject's data.

Formal model comparison
The maximum likelihood associated with each model characterizes how well that model captures patterns in the empirical data. However, comparing models directly in terms of likelihood can be misleading; greater model complexity can allow for tighter fits to the data but can also lead to undesirable levels of overfitting, i.e., the erroneous modeling of random variation in the data. Therefore, we compared the models using the Bayesian Information Criterion (BIC). This measure provides a means for comparing models on the basis of their maximum likelihood fits to the data while correcting for model complexity[8].

Supplementary Information References
1. Gorea, A., & Sagi, D. *Proc. Natl. Acad. Sci. USA* **97**, 12380–12384 (2000).
2. Green, D.M., & Swets J.A. New York: Wiley (1966).
3. Bahrami, B., Lavie, N. & Rees, G. *Curr. Biol.* **17**, 509-513 (2007).
4. Brainard, D.H. *Spat. Vis.* **10**, 433-436 (1997).
5. Watson, A.B., & Pelli, D.G. *Percept. Psychophys.* **33**, 113-120 (1983).
6. Dorfman, D.D., & Alf, E.J. *Math. Psychol.* **6**, 487-496 (1969).
7. Kirkpatrick, S., Gelatt, C.D., Jr, & Vecchi, M.P. *Science* **220**, 671-680 (1983).
8. Burnham, K.P., & Anderson, D.R. New York, NY: Springer (2002).