Filling-in and suppression of visual perception from context — a Bayesian account of perceptual biases by contextual influences

Published in 2008 in PLoS Computational Biology, 4(2): e14 doi:10.1371/journal.pcbi.0040014

Li Zhaoping[1] & Li Jingling[2]

[1]Department of Computer Science, University College London, UK, email: z.li@ucl.ac.uk

[2]Graduate Institute of Neural & Cognitive Sciences, China Medical University, Taiwan

Abstract: Visual object recognition and sensitivity to image features are largely influenced by contextual inputs. We study influences by contextual bars on the bias to perceive or infer the presence of a target bar, rather than on the sensitivity to image features. Human observers judged from a briefly presented stimulus whether a target bar of a known orientation and shape is present at the center of a display, given a weak or missing input contrast at the target location with or without a context of other bars. Observers are more likely to perceive a target when the context has a weaker rather than stronger contrast. When the context can perceptually group well with the would-be target, weak contrast contextual bars bias the observers to perceive a target relative to the condition without contexts, as if to fill in the target. Meanwhile, high contrast contextual bars, regardless of whether they groups well with the target, bias the observers to perceive no target. A Bayesian model of visual inference is shown to account for the data well, illustrating that the context influences the perception in two ways: (1) biasing observers' prior belief that a target should be present according to visual grouping principles, and (2) biasing observers' internal model of the likely input contrasts caused by a target bar. According to this model, our data suggest that the context does not influence the perceived target contrast despite its influence on the bias to perceive the target's presence, thereby suggesting that cortical areas beyond the primary visual cortex are responsible for the visual inferences.

Non-technical summary

We study how visual perception of a target bar can be biased by contextual bars in the image, and how a Bayesian model of object inference can account for the data. Human observers are more likely to perceive a target bar when the contextual contrast, i.e., the luminance difference between the contextual bars and background, is weaker rather than stronger. Relative to the situation without the context, they are biased to perceive the target in a context of weak contrast when the target can perceptually group well with the context, as if the context fills in the target. Meanwhile, they are biased not to perceive the target in a context of strong contrast, as if the context suppresses the perception, regardless of whether it could perceptually group well with the would-be target. The Bayesian model illustrates that the context influences the perception by biasing (1) observers' prior belief that a target should be present and (2) observers' internal model of the likely input contrasts from a target bar. Our data suggest that brain areas beyond the primary visual cortex along the visual pathway are responsible for inferring object causes for input images.

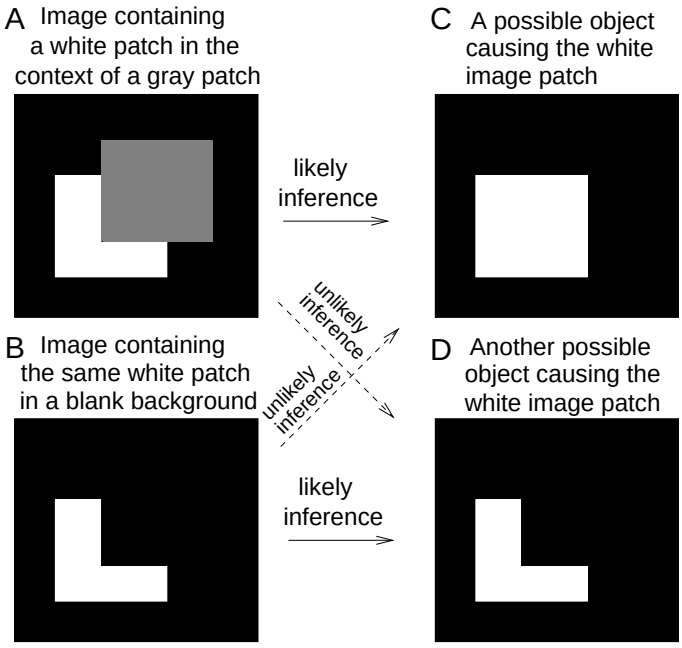

Figure 1: Demonstration of inferences of objects from images. A and B show two images containing the same white patch, C and D show the two possible inferred objects in the scene causing this white patch. The inferred causes for any particular input image patch is not unique, although some inferences are more likely than others. The difference in the most likely inferred object for the same image patch in A and B demonstrates that inference could be greatly influenced by the image context.

1 Introduction

1.1 Background

Visual inputs are first represented in early visual stages such as retina and the primary visual cortex (V1), such that input features such as local color, orientation, luminance contrast, and spatial scale of image patches are encoded by the activities of retinal and V1 neurons with various input sensitivities. The neural representation of inputs is then used by the brain to infer the possible objects in the three-dimensional scene causing the two dimensional input images. For instance, from V1's responses to the luminance edges in Fig. (1A), the brain could infer a white square surface behind a gray square surface, likely employing cortical area V2 where neurons tuned to surface border ownerships signal which of the possible object surfaces is likely responsible for each luminance edge[1, 2]. Information about the object causes are only ambiguously available, or even apparently missing, in the two-dimensional images. As vision is an under-constrained or ill-posed problem, the possible objects causing a given image are not unique. For instance, the white L-shaped image patch in Fig. (1A) is likely caused by a white square surface behind the gray one in the 3-D world; but it is not impossible, though less likely, that an L-shaped surface is the cause. Nevertheless,

perception is rarely ambiguous, typically revealing only (the most likely) one cause at any time given an input. Here, perception is defined as the result of revealing a cause to visual awareness, while inference is the process of assigning a probability to each cause. As both perception and inference are assessed operationally by the same observer reports, the two words are often used interchangably in this paper. It is difficult to state the veridicality of the perception objectively. For instance, a substantial part of the white square surface (in the 3-D world) is not recorded in the two-dimensional input image, and would be non-veridical in terms of image pixel values rather than the 3-D world.

Visual inference from any part of the input is often influenced by the contextual input. For instance, the more likely cause for the white patch in Fig. (1A) or Fig. (1B) is the square or L-shaped surface respectively, due to the presence or absence of the contextual gray patch. The speed and accuracy to recognize an object, e.g., a sewing machine, significantly depend on, e.g., whether it is in an indoor or outdoor scene[3]; and the color appearance of an image patch depend on the surrounding patches[4]. This is unsurprising since the missing or ambiguous information, e.g., the occluded part of a face or the reflectance of a surface, can only be filled in or deduced from the context through the statistical knowledge about visual scenes, e.g., the correlations between neighboring inputs. Contextual influences are also present in the input encoding. For instance, the sensitivity of a V1 neuron to an input bar can be increased by contextual bars (outside the receptive field of the neuron) aligned with it[5, 6, 7], and this colinear facilitation has been manifested in human sensitivity to detect a small bar or gabor (or grating) patch[8, 9, 10, 11, 12, 13].

We are interested in contextual influences in inference of objects from images, focusing in this paper on the perception in the spatial context of other inputs. Most previous studies on influences by spatial context used quite complex inputs such as photographs of everyday scenes[14, 15], demonstrating very interesting phenomena[16]. However, these complex inputs are difficult to manipulate systematically, and the complex spatial relationships between image features[15] are difficult to describe and model in an intuitive and meaningful way, unless when the exact spatial relationship is not essential such as when inferring surface color appearance[17]. This study uses stimuli that are easy to manipulate and describe. They are composed of several bars, like those used in probing contextual influences on input sensitivity[8, 9, 10, 12, 13].

The previous studies used the stimuli of bars to probe input sensitivities by the two alternative forced choice (2AFC) design. In contrast, we probe perceptual biases by a yes-no design. In each trial of the 2AFC design, two brief intervals of the stimuli are presented, both intervals contain the same contextual input but only one contains the target, and the observer has to answer which interval contains the target. The input sensitivity is inversely linked with the minimum target input (contrast) necessary to enable about 80% of the responses by the observers to be correct. It has long been known[18] that measurements from the 2AFC tasks remove the effect of any perceptual or response bias (e.g., on whether the target bar is present), whether the bias arises from the contextual inputs or other factors. In each trial of a yes-no task, after only one stimulus presentation interval, observers have to answer 'yes" or "no" regarding whether they perceive a target bar, i.e., whether the target rather than noise is the inferred cause of the luminance profile at the would-be target location in the input image. Whether the answer is veridical according to the input images is not

the issue, rather, we assess whether the observer *perceives or infers* the target bar, even if its contrast is missing in the input image. This yes-no task thus assesses the bias (to respond "yes") in inferring the target object. One particular bias is filling-in, which we define as a behavioral indication of a target *object* (by responding "yes") when there is no input contrast at the corresponding image location. Note that filling-in here is *not* defined as (mentally) painting-in a *luminance contrast* at the image location corresponding to the target object when the input contrast is zero. Analogously, amodel perceptual completion of the occluded square (in Fig. (1A)) is achieved without seeing any contrast at the image location for the occluded part of the square.

We report in this paper that our study, using the bar stimuli and the yes-no task, revealed how visual contexts influence the perception of the target bar through a Bayesian inference and decision process. In particular, quite unexpectedly from the finding of colinear facilitation of input sensitivities revealed neurophysiologically and behaviorally (by the 2AFC task), we found that weaker colinear contexts induce stronger biases to fill-in the missing target. In the framework of a model of the Bayesian process, our data suggest that contextual facilitation or suppression of input sensitivities plays no role in the inference probed by our task, and hence the neural substrate responsible for this inference is more likely beyond V1. In the rest of the Introduction, we formulate the Bayesian model applied to our yes-no task. The Result section then presents our experiments probing the contextual influences in human inference behavior and the fit of our data by the Bayesian model. The Discussion section will summarize the findings with discussions.

1.2 The Bayesian model of contextual influence on visual inference from simple bar stimuli

1.2.1 The formulation

The Bayesian inference and decision process applied to our task is formulated as follows[18, 19]. Let a stimulus pattern contain input contrast C_t and C_c for the target and contextual bars respectively, evoking neural responses x_t and x_c respectively in the early visual stages. When the target is absent in the image, $C_t = 0$. For presentation simplicity without loss of generality, the target and context are assumed as sufficiently far apart spatially to evoke dissociable responses. The brain infers from x_t whether the target is present, i.e., whether x_t is caused by the target bar or noise, by assigning a probability $P(yes|x_t)$ that a target is present given response x_t. By Bayesian theorem, $P(yes|x_t) \propto P_{x_c}(x_t|yes)P_{x_c}(yes)$, where $P_{x_c}(x_t|yes)$ is the probability, by the brain's internal model, of response x_t to a target, and $P_{x_c}(yes)$ is the prior probability, believed by the brain, that a target should be present. Hence, $P(yes|x_t)$ is the posterior probability in the Bayesian terminology. Note that $P_{x_c}(x_t|yes)$ is not a typical likelihood term in Bayesian terminology in which the likelihood typically means the conditional probability of neural response x_t if the experimenter presented a target — instead, $P_{x_c}(x_t|yes)$ is what the brain thinks the probability of response x_t should be when the brain assumes that x_t is caused by a target, whether or not the experimenter actually presented the target. The subscript x_c in $P_{x_c}(x_t|yes)$ and $P_{x_c}(yes)$ indicates that both could be influenced (or parameterized) by the response x_c to the context. To minimize the mean response error (assumed as the loss function in the decision), the observer's optimal response to the question

"is the target present?" is "yes" when $P(yes|x_t) > 0.5$ and "no" otherwise. With input and neural noise, the neural responses x_t (and x_c), and consequently $P(yes|x_t)$ and the observer's response, can vary from one trial to another given a fixed input presentation. Averaged over many trials of a given input image, one can measure the probability $P(yes|C_t)$ of response "yes" given a target contrast C_t (and context). We can phenomenologically call $P(yes|C_t)$ the posterior, as the brain's inferred probability of a target being present given the input contrast C_t. It is the counterpart or the manifestation of $P(yes|x_t)$, internal to the brain and inaccessible to our behavioral measurements. The Appendix gives a detailed formulation to arrive analogously at the phenomenological internal model $P(C_t|yes)$ and phenomenological prior $P(yes)$, the counterparts of $P_{x_c}(x_t|yes)$ and $P_{x_c}(yes)$ respectively. For simplicity in the main text, we use this phenomenological language to present the rest of our formulation of the inference process, and omit the details of the decision process (of choosing to respond "yes" or "no" given $P(yes|x_t)$) unless it is necessary (e.g., in the Discussion section). To avoid notational clutter, different probabilities, e.g., $P(yes)$ and $P(C_t|yes)$, are simply denoted by the differences in the variables, with no or minimum notations for the parameter dependences.

In the Bayesian model, the inferred probability $P(yes|C_t)$ that C_t is caused by a target bar arise from weighing the two probabilities: one is the probability $P(yes)P(C_t|yes)$ that C_t could arise from a target, the other is the probability $P(no)P(C_t|no)$ that C_t could arise from "no target" or noise. Here $P(yes)$ and $P(no) = 1 - P(yes)$ are the prior probabilities, assumed by the brain, of a target as present and absent respectively; and $P(C_t|yes)$ and $P(C_t|no)$ are the brain's internal models of the probabilities of having input contrast C_t at the would-be target location when the brain assumes the target is present or absent respectively. Hence

$$P(yes|C_t) = \frac{P(C_t|yes)P(yes)}{P(C_t|yes)P(yes) + P(C_t|no)P(no)} \tag{1}$$

Note that $P(yes)$, $P(no)$, $P(C_t|yes)$, and $P(C_t|no)$ are the internal belief or models in the observer's brain. In particular, $P(yes)$ is *not* the probability that the experimenter actually presented a target bar at the target location, nor is $P(C_t|yes)$ the probability that a contrast C_t is presented at the target location by the experimenter, the "yes" in $P(C_t|yes)$ refers to the brain's assumed condition of a target present rather than the actual presence of a target placed by the experimenter. Throughout the paper, "yes" and "no" always refer to the observer's responses or internal variables in his/her brain rather than the experimenter's stimulus presentation.

Both $P(yes)$ and $P(C_t|yes)$ are subject to observer's biases which can be influenced by the context, as illustrated in Figure (2). If one occluded from view the target but not the contextual bars, the prior $P(yes)$ is the observer's expected probability that the target is present behind the occluder. So $P(yes)$ is higher in a colinear context which is seen as more likely to group with target. The context also influences $P(C_t|yes)$ by making observers expect that the target and contextual bars should have similar contrasts, i.e., the probability $P(C_t|yes)$ of the target contrast C_t should peak around $C_t = C_c$ (see Figure (2B)). We thus model

$$P(C_t|yes) = \frac{\exp(-|C_t - C_c|/\sigma_y)}{N_y}, \tag{2}$$

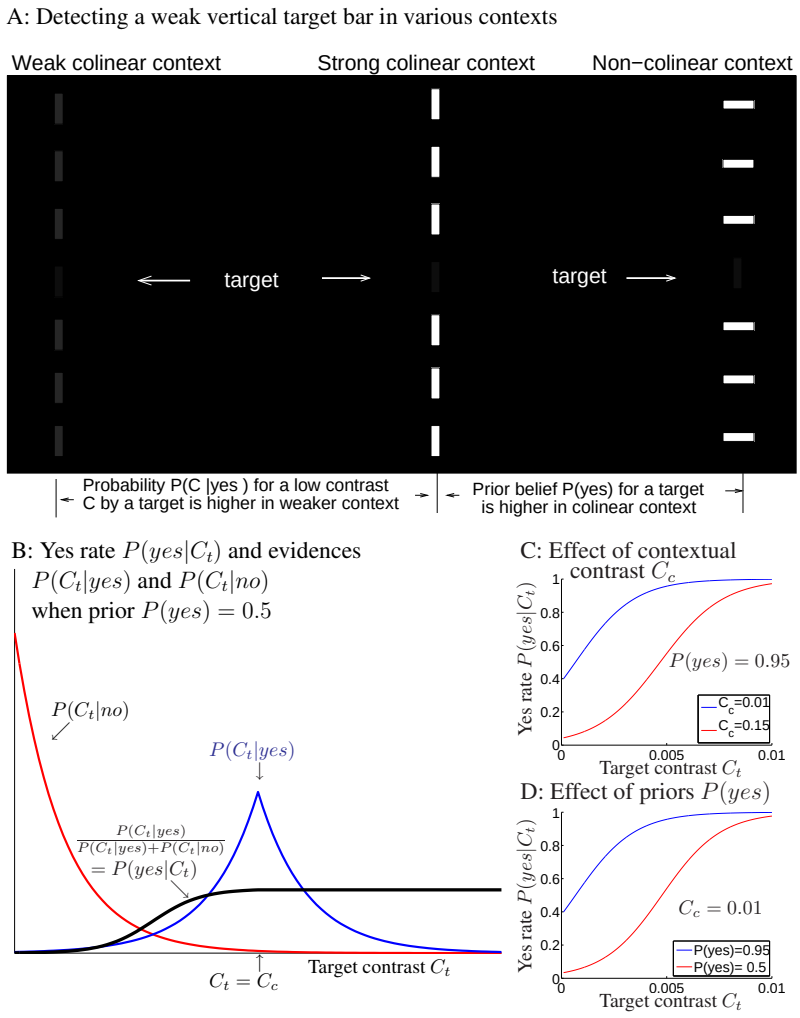

Figure 2: Bayesian inference for target perception. A: schematics of perceiving a weak vertical target bar in three different contexts. Colinear contexts give a higher prior belief $P(yes)$ of the target present, as it could be grouped with the context. Higher contextual contrast C_c makes a low contrast input C_t at the would-be target location seem less likely to be caused by a target rather than noise, since observers expect a target to evoke a contrast similar to C_c, i.e., $P(C_t|yes)$ peaks at $C_t \approx C_c$, and $P(C_t|yes) \approx 0$ if $C_t \ll C_c$, see B. B: the probability $P(yes|C_t)$ of "yes" response depends on the ratio between the evidences $P(C_t|yes)$ and $P(C_t|no)$ for target present and absent respectively, when the prior belief $P(yes) = 0.5$ is unbiased. This ratio should be multiplied by $P(yes)/(1 - P(yes))$ in general. Note that probability distributions $P(C_t|yes)$ and $P(C_t|no)$ peak at $C_t = C_c$ and $C_t = 0$ respectively. C and D: effects of the contextual contrast C_c (in C) and of the prior $P(yes)$ (in D) by the Bayesian model. In C and D, all curves have model parameters $k = 2$, and $\sigma_n = 0.0015$, the two red curves are identical, with $P(yes) = 0.95$ and $C_c = 0.01$. Comparing C and D, a higher contextual contrast C_c has a similar effect as a lower prior $P(yes)$.

where σ_y models the uncertainty about the target contrast, $N_y = \sigma_y[2 - \exp(-C_c/\sigma_y) - \exp(-(1 - C_c)/\sigma_y)]$ is the normalization constant for the probability distribution on the contrast range $0 \leq C_t \leq 1$. It is reasonable to assume (see Appendix for justifications) that σ_y is proportional to C_c with a Weber-like scale factor k,

$$\sigma_y = k \cdot C_c. \tag{3}$$

Without the context $P(C_t|yes)$ is assumed (its exact form does not matter, as it is never fitted to the data) to become $P(C_t|yes) \propto \exp(-C_t/\sigma_0)$ with a contrast uncertainty σ_0. The brain also assumes that input contrast C_t caused by noise or other non-target factors to be near zero, hence,

$$P(C_t|no) = \frac{\exp(-C_t/\sigma_n)}{N_n}, \tag{4}$$

where $N_n = \sigma_n[1 - \exp(-\sigma_n^{-1})]$, with contrast uncertainty σ_n determined by the observer's internal model of the noise. From equations (1-4), we see that three parameters: $P(yes)$, k, and σ_n can completely model $P(yes|C_t)$ for all C_c and C_t, given a contextual configuration which determines $P(yes)$.

1.2.2 The elaborations

One may think of $P(C_t|yes)$ and $P(C_t|no)$ as evidences for a target present and absent, respectively, and the observer arrives at his response probability $P(yes|C_t)$ by combining the evidences with his prior belief $P(yes)$ and $P(no)$. Both the priors and the evidences are influenced by the context — the prior $P(yes)$ by the contextual configuration while the evidence $P(C_t|yes)$ by the resemblance between the contextual contrast C_c and the input contrast C_t. In general, one could model the evidence $P(C_t|yes)$ and prior $P(yes)$ such that each could be affected by *both* the configuration *and* the contrast of the context. Insufficient motivation for such a generality, which would nevertheless require additional model parameters, justifies eliminating it by Occam's razor.

Fig. (2C) illustrates that a higher contextual contrast C_c gives a lower $P(yes|C_t)$ or suppresses the perception of a target with small C_t, since it makes the low contrast C_t seem as unlikely caused by a target rather than noise. This is because, when the context is clearly visible while the target is barely visible, $C_t < C_c$ (as is always the case in our experiment), the evidence $P(C_t|yes) = \exp[-(C_c - C_t)/(kC_c)]/N_y$ decreases with increasing C_c. In detail, if context one and context two have the same configuration but different contrasts C_{c1} and C_{c2} such that $C_{c1} > C_{c2} > C_t$, let P_{c1} and P_{c2} denote the probability $P(C_t|yes)$ under C_{c1} and C_{c2} respectively, then, $P_{c2}/P_{c1} \propto \exp[(C_t/k)(\frac{1}{C_{c2}} - \frac{1}{C_{c1}})] \geq 1$ (provided that the normalization constant N_y for C_{c1} is larger than that for C_{c2}, which is indeed the case for us, as shown in the Apprendix). Meanwhile (see Fig. (2D)), given a contextual contrast C_c (and thus the evidence $P(C_t|yes)$), one is more likely to expect a target in the colinear than non-colinear context since the prior belief $P(yes)$ is higher in the colinear context.

Fig. (2B) illustrates that in some ranges of input contrast C_t, the evidences $P(C_t|yes)$ and $P(C_t|no)$ for and against a target's presence, respectively, are very different from each other, i.e., $P(C_t|yes)/P(C_t|no) \to \infty$ or 0. In such a case, the evidences are unambiguous, diminishing the effect of a prior $P(yes)$, making the responses (with probability $P(yes|C_t)$) also unambiguous. This

happens over a large range of small C_t when a stronger contextual contrast C_c pulls the distributions $P(C_t|yes)$ and $P(C_t|no)$ apart from each other. When C_c is sufficiently low, there is a sizable range of low input contrast C_t in which the evidences $P(C_t|yes)$ and $P(C_t|no)$ for and against a target are comparable, i.e., the evidences are ambiguous, giving the prior $P(yes)$ the power to sway the response probability $P(yes|C_t)$.

Filling-in, which occurs when $C_t = 0$ but $P(yes|C_t)$ is substantial, is an example when the prior sways the response. It happens particularly when the noise level σ_n is high, such that a zero input contrast C_t could be caused by the target *or* the noise, i.e., $P(C_t = 0|yes)$ is non-negligible compared to $P(C_t = 0|no)$. The observer's "yes" response when $C_t = 0$ is analogous to perceiving a white square in Fig. (1A) without perceiving any luminance contrast at the image location for the occluded corner of the square. For the partially occluded square, perception attributes the missing luminance to the occluder. For the filled-in target bar, perception attributes the zero contrast $C_t = 0$ to input or neural noise (such as the noise in the photoreceptors or V1 neurons), which causes input contrasts and/or brain responses to fluctuate away from their supposed levels in the noise-free situation. Hence, a "yes" response to zero target contrast, the result of a decision based on a perception (even if vaguely) of the target, is no less veridical than the perception of the partially occluded square. Analogously, one may perceive no target even under non-zero input contrast C_t, when the evidence $P(C_t|yes)$ for a target is insufficient and C_t is attributed to, or explained away by, noise, depressing the posterior probability $P(yes|C_t)$.

The Bayesian inference described above predicts in particular: (1) a weak context encourages filling-in of the visual target object when it is consistent or easily grouped with the target, i.e., $P(yes)$ is large; (2) a sufficiently strong context can suppress the perception of a weak target since the strong context bias the observer to presume a weak input contrast C_t as caused by noise rather than a target; and (3) the prior belief $P(yes)$ can be influenced by the spatial configuration of the context in a way that is consistent with the statistical properties of visual inputs. We report experiments confirming the predictions next.

2 Results

In the experiments, human observers were asked to answer whether or not they perceive the target by pressing a button. They were informed that the target when present was a nearly visible vertical bar at the center of the fixation array, and that they should make their judgments according to the target alone regardless of the context. We only used naive observers to minimize any systematic bias not related to the contextual stimuli. In each trial, the particular target and contextual (contrast and configuration) condition was unpredictably chosen among all conditions within an experiment.

2.1 Experiment 1: weaker contexts give higher yes rates $P(yes|C_t)$

In experiment 1, the context has 10 colinear bars on each side of the target bar (Fig. (2A)), and its contrast can be one of $C_c = 0$, 0.01, 0.05, and 0.4, with $C_c = 0$ for the no context baseline

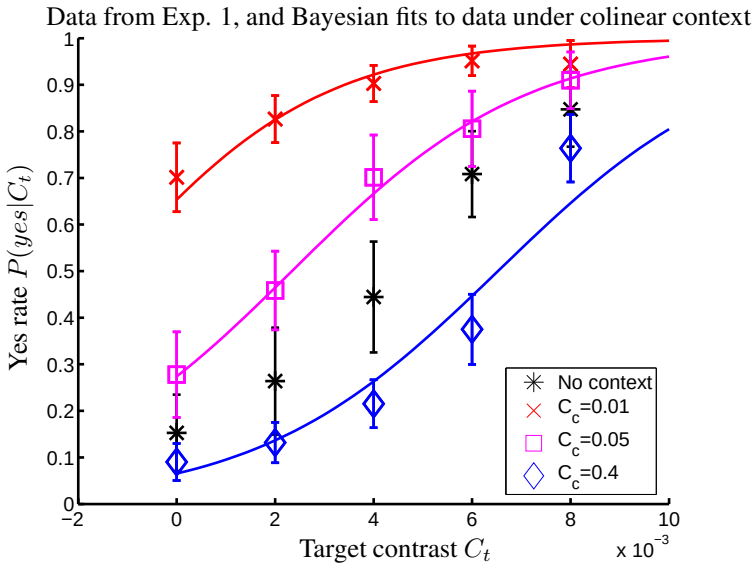

Figure 3: Results from experiment 1, where the colinear context resembles the two left ones in Fig. (2A). The data points are the mean over six observers, and the error bars indicate the standard errors of the means (SEMs). On average and relative to the no context condition, the weaker colinear contexts $C_c = 0.01$ and $C_c = 0.05$ raised the yes rates by CFI = $(38 \pm 8)\%$ and $(15 \pm 8)\%$, respectively, whereas the stronger context $C_c = 0.4$ lowered it by -CFI = $(17 \pm 8)\%$. The colored curves are Bayesian fits to data of the corresponding color, no fit is done for data without context. The root mean square normalized fitting error RMSNFE = 0.66 in the unit of SEM. The fitted parameters (and their 95% confidential intervals) are $k = 1.9(0.6, 3.2)$, $\sigma_n = 0.0025(0.0020, 0.0029)$, and $P(yes) = 0.972(0.967, 0.978)$.

condition. This is to investigate whether weaker and stronger contexts do give higher and lower yes rates $P(yes|C_t)$ respectively as predicted. Here contrast is defined by Michelson contrast $C = (L_{max} - L_{min})/(L_{max} + L_{min})$ where L_{max} is the luminance of the bar and L_{min} that of background. Each bar is a rectangle of $0.9^o \times 0.165^o$ in size, and the centers of the neighboring bars were 1.15^o apart. The possible target contrast $C_t = 0, 0.002, 0.004, 0.006$, and 0.008 span a range from below to somewhat above the typical human contrast detection threshold without context. Each test image was presented for 24 trials for each observer.

We found that (Fig. 3), compared to the yes rates under no context, the mean yes rates averaged over six observers are higher under low contextual contrast $C_c \leq 0.05$ and lower under higher contextual contrast $C_c = 0.4$, for any target contrast C_t. We define a contextual facilitation index (CFI) as the average increase in the yes rate in a particular context (relative to no context), specifically

$$\text{CFI} \equiv \text{Mean}_{C_t}[P(yes|C_t, \text{a given context}) - P(yes|C_t, \text{without context})] \tag{5}$$

where $\text{Mean}_{C_t}(x) \equiv [\sum_{C_t} x]/[\sum_{C_t} 1]$ stands for the average of x over C_t. The weakest context $C_c = 0.01$ raises the yes rate by CFI = 0.38 ± 0.08, and the intermediate context $C_c = 0.05$ by CFI

= 0.15 ± 0.08. In contrast, the strongest context $C_c = 0.4$ *lowers* the yes rate by |CFI| $= 0.17 \pm 0.08$. Averaged over C_t, the observers were more than twice as likely to perceive a target in the weakest than in the strongest context.

The mean yes rates with the context are $(86 \pm 4)\%$, $(63 \pm 6)\%$, and $(32 \pm 5)\%$ respectively for $C_c = 0.01$, 0.05 and 0.4 and $(48 \pm 9)\%$ without the context. However, the mean yes rate over trials of all target and contextual conditions is $(57 \pm 5.5)\%$, suggesting that observers have an internal, stimulus unrelated, prior to roughly equalize their total numbers of "yes" and "no" responses, even though we did not give them any indication of the expected rate of "yes" responses. If the experiment had only one contextual (contrast and configuration) condition, this internal prior could at least partly overwrite the prior caused by the context. Hence, interleaving different contextual conditions within a session helps to manifest and differentiate perceptual biases caused by different contexts.

The adequacy of the Bayesian model is demonstrated by its reasonable fit to the data from the three non-zero contextual contrast conditions, using only three parameters k, σ_n, and $P(yes)$. Let $P_{data}(yes|C_t)$ and $P_{fitted}(yes|C_t)$ be the measured (mean) and fitted yes rates, and $\delta P_{data}(yes|C_t)$ the (SEM) error of $P_{data}(yes|C_t)$, and $E \equiv P_{data}(yes|C_t) - P_{fitted}(yes|C_t)$ the fitting error. For each data point i denoting a particular contextual and target condition, we denote the fitting error and the SEM error as E_i and δ_i respectively. The quality of the Bayesian fit for a total of N data points can be quantified by the root mean squared normalized fitting error defined as

$$\text{RMSNFE} = [(\sum_{i=1}^{N} E_i^2/\delta_i^2)/N]^{1/2}, \tag{6}$$

which indicates the fitting error in the units of the SEM errors of the mean yes rates. When RMSNFE< 1, for instance, the fitted curve is within the size of the error bars from the measured data for typical data points. The fitting finds the optimal set of Bayesian model parameters k, σ_n and $P(yes)$ that minimizes this RMSNFE. Our fit to a total of $N = 3 \times 5$ data points for the 3 yes rate curves gives RMSNFE $= 0.66$. Note that, a psychometric function parameterized by two or more parameters can typically fit a single yes rate curve (which in our case contains 5 data points). For instance, a logistic function $P(yes|C_t) = 1/(1 + \exp((\alpha - C_t)/\beta))$ with two parameters, α and β, could also reasonably fit a yes rate curve in our data. However, three logistic functions or a total of six parameters would be needed to fit three yes rate curves. Hence, fitting our data for three yes rate curves within the error bar by the Bayesian model, using only a total of three parameters, reflects the adequacy of the Bayesian account.

Note that fitting the yes rate data for the no context condition by the Bayesian model would require two additional parameters, σ_0 and the prior probability $P_{no\ context}(yes)$ under no context, as many as needed by the logistic fit. Hence, fitting this curve well by the Bayesian model adds no additional strength to the Bayesian account. In fact, since the parameter σ_n is already determined from fitting the three yes curves for the colinear context, the two additional Bayesian parameters σ_0 and $P_{no\ context}(yes)$ are under determined (i.e., many different choices of σ_0 and $P_{no\ context}(yes)$ would give roughly equally good fits) for a curve that needs only two essential parameters. Thus we display these data as they are without any model fitting.

The higher yes rates under weaker contextual contrasts C_c are not expected from the assumption or expectation that neurons responding to the colinear context should increase the neural response to the target as if the target has an effective contrast $C_t^{effective}$ higher than the actual input contrast C_t. If colinear facilitation did make $C_t^{effective} = C_t + \Delta C_t$, then the change ΔC_t should depend on the contextual contrast C_c by some function as $\Delta C_t = f(C_c)$ such that $f(0) = 0$. Then, our Bayesian formulation should replace each C_t in the right hand side of equation (1) by $C_t^{effective}$. To the first order (linear) approximation, $\Delta C_t \approx \gamma C_c$, where γ is the coefficient of facilitation. We can then repeat our Bayesian fit with now an additional model parameter γ. As expected, this gives a negligible fitted $\gamma = -0.5 \times 10^{-6} \approx 0$, giving $|\Delta C_t| < 10^{-5}$ for $C_c \leq 0.4$. Hence, no colinear facilitation or suppression of input sensitivities is needed to account for our data, or that our data do not indicate that colinear influence could change the effective contrast of the input.

2.2 Experiment 2: colinear and orthogonal contexts

Experiment 2 was based on Fig. (2A), to test that different spatial configurations, one colinear and one orthogonal, of the context can give rise to different prior probabilities $P(yes)$ according to observers' belief. The colinear context was the same as that in experiment 1, while the orthogonal context differs from the colinear one only by the orientation of the contextual bars. The contextual contrast used were $C_c = 0.01$ and 0.4, with another $C_c = 0$ serving as the no context baseline. Five observers participated in this experiment, each took 20 trials for each condition of a given C_t, C_c, and spatial configuration of the context.

Fig. (4) shows the results. Regardless of the contextual configuration, the yes rate is higher when the contextual contrast C_c is lower, CFI($C_c = 0.01$)−CFI($C_c = 0.4$) $>\approx 0.4$, and a sufficiently high C_c gives negative CFI, biasing the observers to respond "no". For every contextual contrast C_c, the colinear context gives a higher yes rate than the orthogonal one, CFI(colinear) - CFI(orthogonal)$>\approx 0.23$. At low contextual contrast C_c, the colinear context biases the response to "yes" (CFI> 0), while the orthogonal context gives no significant bias. These findings are consistent with our qualitative arguments in Fig. (2).

The data can be fitted by the Bayesian model for the four yes rate curves (two configurations $\times$ two contextual contrasts) using only four parameters: k, σ_n, and the prior probabilities $P(yes)_{colinear}$ and $P(yes)_{orthogonal}$, with each data point typically about one error bar size away from the model fit. As expected, $P(yes)_{colinear} > P(yes)_{orthogonal}$ (Fig. (4E)). However, both $P(yes)_{colinear}$ and $P(yes)_{orthogonal}$ are quite high. This we believe is the net result of combining two factors, one is the observers' internal prior to reach roughly equal numbers of "yes" and "no" responses, and the other is the contextual dependent priors from the statistical knowledge of the natural visual environment. Indeed, the average yes rate (over all trials and observers) is $(57 \pm 2)\%$. The difference between the fitted $P(yes)_{colinear}$ and $P(yes)_{orthogonal}$ reflects the difference between the natural priors that has survived observers' internal prior imposed by the unnatural laboratory experiment.

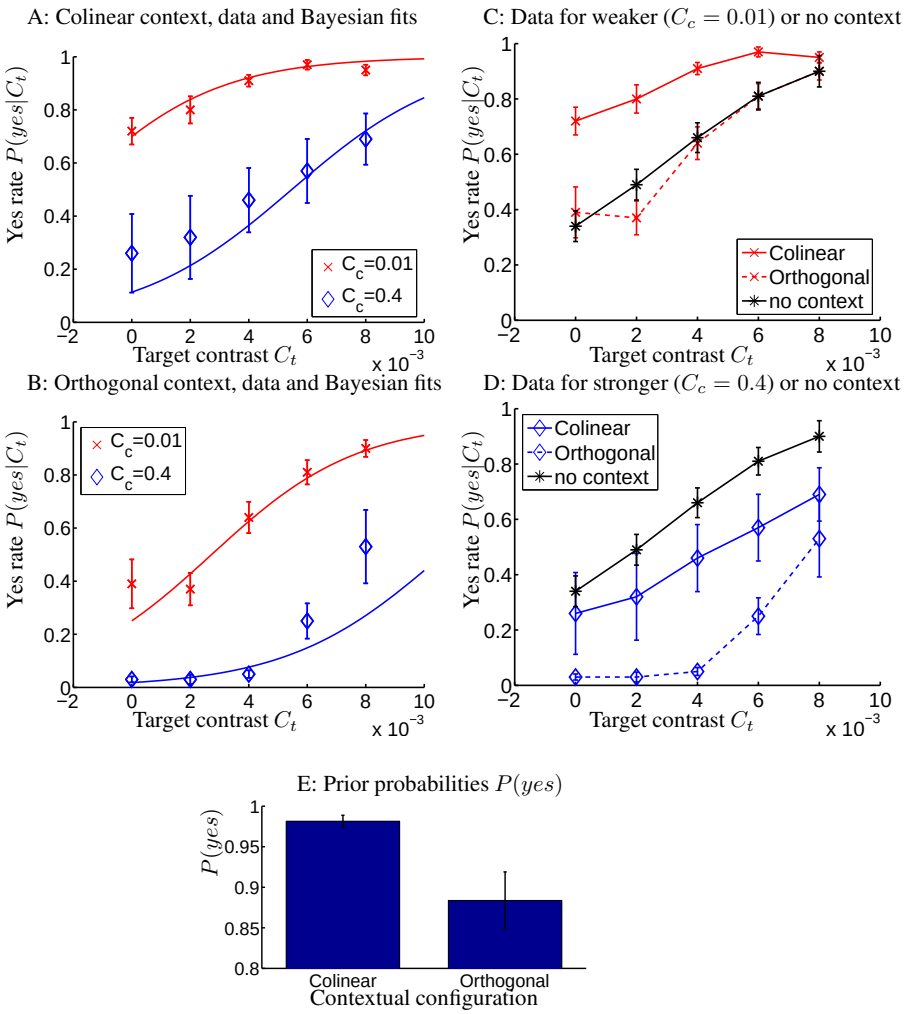

Figure 4: Results from experiment 2 averaged over five observers. A & B : Yes rates under colinear and orthogonal context (schematically like Fig. (2A)) respectively. The curves are the Bayesian fits. The four Bayesian parameters (and their 95% confidence intervals) are $k = 3.8(1.8, 5.8)$, $\sigma_n = 0.0027(0.0021, 0.0033)$, $P(yes)_{colinear} = 0.982(0.974, 0.989)$, and $P(yes)_{orthogonal} = 0.88(0.85, 0.92)$, giving a fitting quality of RMSNFE = 1.0. C & D: Yes rates under different contextual contrast $C_c = 0.01$ and $C_c = 0.4$ respectively, together with those under no context. For colinear context CFI = 0.23 ± 0.05 and -0.18 ± 0.15 for $C_c = 0.01$ and 0.4 respectively; for orthogonal context CFI = -0.018 ± 0.06 and -0.46 ± 0.055 for $C_c = 0.01$ and 0.4 respectively. E: priors $P(yes)$ for the two contextual configurations. The error bars denote SEMs in A-D, and 95% confidence intervals in E.

2.3 Experiment 3: different configurations of colinear context

Experiment 3 shows that even subtle differences in contextual configuration can manifest in different biases in inferences in ways consistent with the Bayesian account. It is like Experiment 2, but

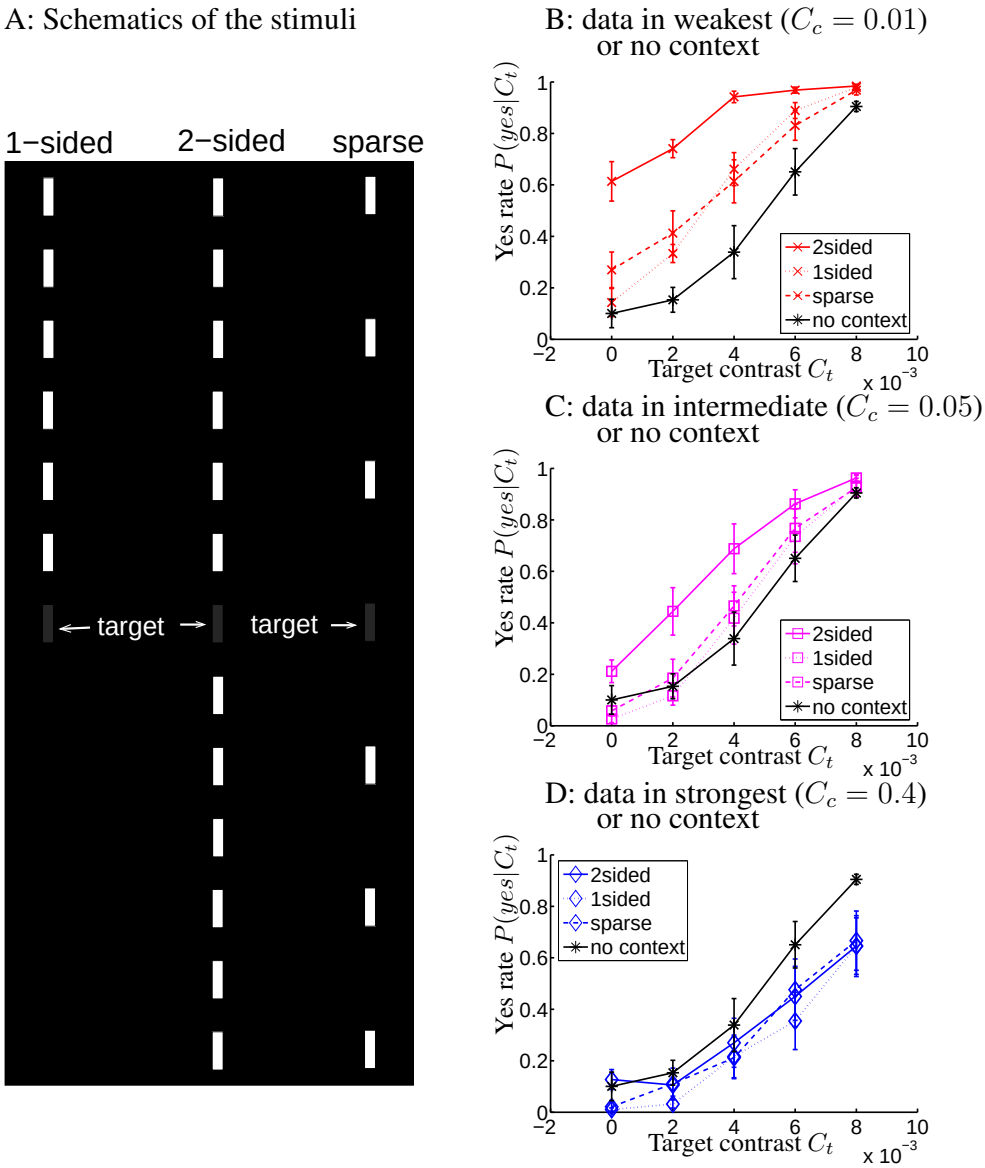

Figure 5: The schematics of the stimuli (A) and the yes rates (with SEM error bars) averaged over seven observers (B-D) from Experiment 3. The 2-sided context gives higher yes rates than other contexts for $C_c = 0.01$ (B), $C_c = 0.05$ (C), but not significantly for $C_c = 0.4$ (D) when yes rates are all depressed relative to those under no context. The yes rates given a contextual configuration decreases with increasing C_c. Error bars indicate SEM. CFI under the 2-sided, 1-sided, and sparse contexts are respectively: CFI = 0.42 ± 0.06, 0.17 ± 0.04, and 0.19 ± 0.04 for $C_c = 0.01$, CFI = 0.204 ± 0.06, 0.016 ± 0.07, and 0.05 ± 0.06 for $C_c = 0.05$, and CFI = -0.11 ± 0.07, -0.18 ± 0.05, and -0.13 ± 0.06 for $C_c = 0.4$.

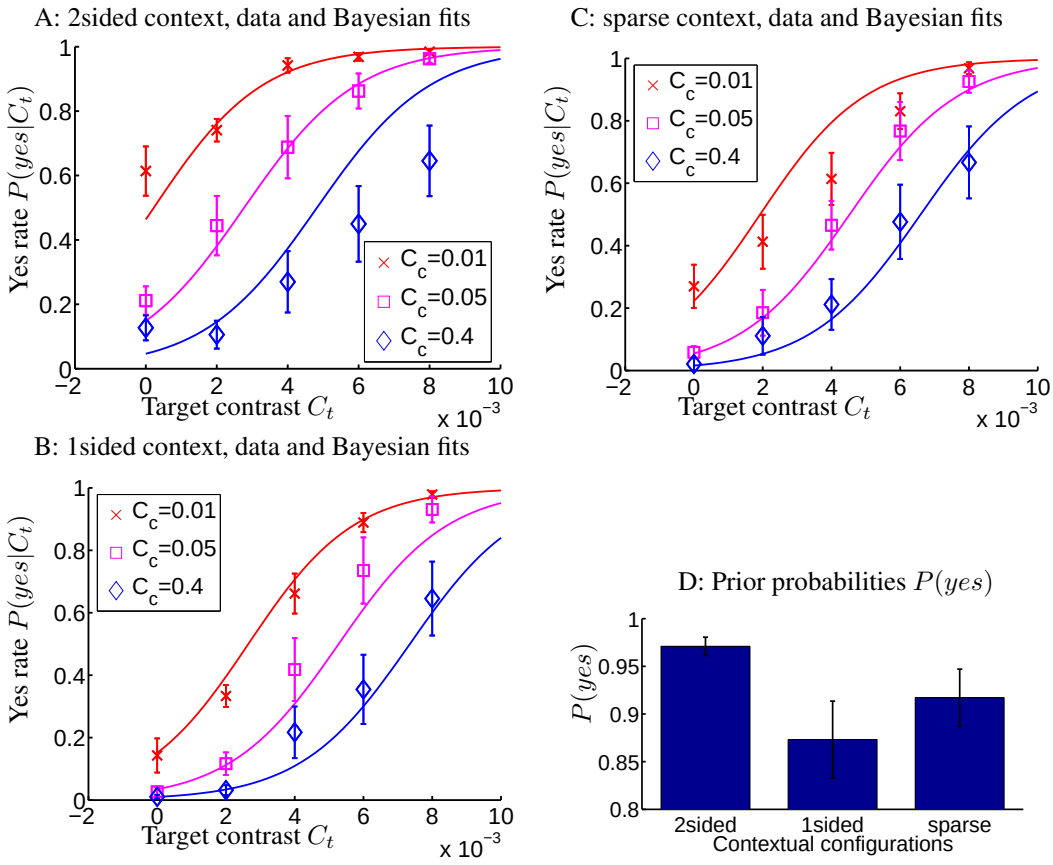

Figure 6: A-C: Fit to data in experiment 3 by the Bayesian model. The red, magenta, and blue curves and data points indicate respective quantities associated with different contextual contrasts $C_c = 0.01, 0.05, 0.4$ respectively. The fitted Bayesian parameters (and their 95% confidential intervals) are $k = 3.91(1.95, 5.87)$, $\sigma_n = 1.60(1.45, 1.73) \times 10^{-3}$, and $P(yes) = 0.97(0.96, 0.98)$ for the 2-sided, $P(yes) = 0.87(0.83, 0.91)$ for the 1-sided, and $P(yes) = 0.92(0.89, 0.95)$ for the sparse context. RMSNFE = 1.07. D: the priors $P(yes)$ for the three different contextual configurations. The error bars denote SEMs in A-C, and 95% confidence intervals in D.

with three colinear context: one is 2-sided which is the one in exp. 1, removing contextual bars from one end of the target gives the 1-sided context, while removing every alternate contextual bar gives the sparce context, see Fig. (5A). The non-zero contextual contrasts are $C_c = 0.01, 0.05$, and 0.4. Each of the seven new observers took three sessions of data to perform a total of 27 trials for each context condition and C_t.

Fig. (5B-5D) show that, the yes rates in the three contextual configurations are very similar for high contextual contrast $C_c = 0.4$, but the 2-sided context gives the highest yes rates under lower $C_c = 0.01$ and 0.05, having CFI values about 0.2 higher than those in other contexts. This is consistent with the expectation that the 2-sided context should have the highest prior, and that the subtler differences between the configurations are more easily manifested under lower C_c conditions when

observers rely more on the priors for their decisions. Meanwhile, as in Experiment 1-2, yes rates decrease with increasing C_c in all contextual configurations. Fig. (6) demonstrates that the data in the nine yes rate curves for the non-zero contexts in this experiment can be reasonably well fitted by the Bayesian model using only 5 parameters k, σ_n, and the three $P(yes)$ values for the three contextual configurations. The $P(yes)$ for the 2-sided context is indeed the highest, even though, as in experiment 2, the differences between the three $P(yes)$'s must be reduced, by the observers' internal prior, from the true differences between the natural priors.

3 Discussion

3.1 Summary of results

Using simple visual stimuli of bars familiar in psychophysical and physiological studies of *input sensitivities*, our study is one of the first to investigate how visual context bias the *perception* of such visual inputs. In particular, the perception is of the presence or absence of a target bar of a known orientation and shape at a central location given a low or zero input contrast at this location, in the context of other input bar stimuli. We showed that high contrast contextual bars bias the observers to perceive no target bars, as if the context suppresses the perception of the target. Meanwhile, low contrast contextual bars aligned with the target bar bias the observers to perceive a target bar, even when there is zero target contrast in the input image, as if the context fills in the target. This filling-in bias is stronger when the contextual bars have weaker contrasts, and when the target is seen as more likely to group with the context as a straight line.

We show additionally that these findings, unexpected from previous findings of contextual facilitation on input sensitivities, can be accounted for by a Bayesian inference and decision model. The model assumes that the perception results from an inference of the posterior probability $P(yes|C_t) \propto P(yes)P(C_t|yes)$ from the following factors: (1) a context dependent prior belief of probability $P(yes)$ and $P(no) = 1 - P(yes)$ of possible visual events "yes" and "no" regarding the target's presence, (2) a (noisy) observation of visual input (contrast) C_t, and (3) the brain's internal model of the context dependent probability $P(C_t|yes)$ or $P(C_t|no)$ of the C_t that could be caused by a target or noise. A context that can be better grouped with the target leads to a stronger prior belief $P(yes)$ of a target's presence. A weak or even zero input contrast C_t is a more plausible evidence for a target ($P(C_t|yes) \gg 0$) in a weaker contextual contrast C_c, since the target is also expected to have a low contrast. In such a case, since evidence $P(C_t|no)$ for C_t as caused by noise is also non-negligible, the input signal-to-noise is often insufficient to dictate the inference, making the inferred probability $P(yes|C_t)$ easily swayed by the prior $P(yes)$. This leads to filling-in when input contrast $C_t = 0$ but inferred probability $P(yes|C_t)$ for the target is substantial. In contrast, a high contrast of the contextual bars makes a weak input contrast C_t as seem unlikely caused by a target rather than noise, i.e., $P(C_t|yes) \approx 0$, suppressing the perception of target, i.e., $P(yes|C_t) \approx 0$, even with a large prior belief $P(yes)$.

3.2 Relating to previous studies

The filling-in and suppression of the target respectively in our study is not unlike the visual assimilation and contrast respectively in the perception of brightness[20], color[21, 22], tilt[23], or motion direction[24], when the contextual features (brightness, color, tilt, motion) make the target feature appear to shift, respectively, towards or away from the contextual feature. At least in the motion perception, there is also a similar correlation between motion capture versus motion contrast (or induction), analogous to our filling-in versus suppression, and the low versus high signal-to-noise of inputs[24]. In the image encoding process before object inference, there is a similar relationship between the shape of the receptive fields and the signal-to-noise in input — when the input noise is high, the receptive fields of the retinal ganglion cells are large and not spatially opponent, leading to input smoothing which is similar to assimilation; when the input noise is low, the receptive fields have the center-surround spatially opponent shape to enhance input contrast. Such a strategy at the input encoding stage has been understood computationally by efficient coding of visual input information[25, 26].

The findings in higher level vision[3, 14, 15, 27, 28] that consistent context can facilitate or speed up object recognition or attentional guidance is analogous to our finding that contexts that can be more easily grouped with the would-be target is more conducive to filling-in, reflecting an inference based on information redundancy or correlations in natural scenes. Analogous phenomena of perceptual completion from context are also ubiquitous in mid-level vision[29], including the completion of the missing or incomplete information on object surface color[4], and on occluded or unoccluded surface boundaries[30].

Compared with most of the previous studies on the influences by the spatial context, our study uses simpler stimuli that can be more easily or quantitatively manipulated and described. Consequently, we not only model our data using a simple Bayesian inference and decision model, but also use this model to deduce that, at least in inference, the underlying neural mechanisms do not cause contextual facilitation or suppression of input sensitivities observed at the visual encoding stage[6, 31]. Some of the previous studies[4, 14], using more controlled stimuli, have also shown that human inference is like that of an ideal observer in a Bayesian inference. In these studies, the Bayesian inferences were based on the *known or built in statistics* of visual inputs. In comparison, we model a Bayesian influence *using a model of* the visual input statistics, parameterized by $P(yes)$, k, and σ_n, which we show is *consistent with the Gestalt grouping laws* which in turn is presumably based on the actual statistics of natural visual inputs. Furthermore, since the target input was independent of the context in the stimulus presentation by the experimenter, the observers' context-dependent perception of the target suggests that they did not modify their internal belief or statistical model of the visual world by sampling the recent stimulus inputs for the task.

3.3 Discussions of various issues

Context can change sensitivity to input bars (or bar like elements such as gabors) as manifested behaviorally in 2AFC tasks for target detection[8, 9, 10, 11, 12], as if the context effectively changes the input contrast. The primary visual cortex has been argued as the neural substrate for such con-

textual influences[5, 6, 31]. However, in our yes-no task probing the inference process, the context does not shift the perceived input contrast from the veridical one according to our model, suggesting that either the brain areas receiving inputs from V1 can somehow distinguish between input sensitivities and input contrast (see[13, 32] for related findings), or that the yes-no task somehow evokes the brain to turns off the contextual influences on input sensitivites[33, 34]. Hence, the neural substrates responsible for visual inference, in particular for associating neural response x_t with the probability $P(yes|x_t)$ for a target object, may be beyond V1. This is consistent with the physiological finding[35, 36] that V2 rather than V1 is more likely responsible for the illusory contours or disparity capture inferred from the contextual inducers[37], analogous to our filled-in target induced by the context. Also consistent with our finding is the observation[38] that neurons in V2 but not V1 respond to illusory brightness of Cornsweet illusion which manifests the inference of surface (but not image) properties, analogous to the inference of a target object but not contrast features in our task. However, our finding does not preclude the possibility that the inference signals being fed back to V1 from higher cortical areas in subsequent or more advanced processes of inference[39, 40]. Different mechanisms for input discrimination (sensitivity) and object appearance (inference) have also been demonstrated behaviorally in luminance and surface processing[41].

In previous studies of contextual influence on visual inferences, researchers probed perception by asking the observers to report the appearance, e.g., color and motion direction, of the stimuli. Our study may seem different by asking for reports of whether the target is perceived or not, rather than the appearance, e.g., apparent contrast. However, in essence, the question of "whether you perceive the target or not" is not unlike a question "whether the luminance profile at this location appears as if it is caused by a target or by noise", which probes the appearance of the perception evoked by the input at the image location concerned. If we had instead asked for reports of apparent contrast, these reports may or may not directly reflect the process of *inferring* the underlying *surface objects* causing the contrast; rather, they may instead reflect the process of *encoding* the two-dimensional *image* property. In a previous study on color matching[42], observers' responses when asked about the hue and saturation of input showed little color constancy, i.e., the responses did not reflect the underlying surface causes; meanwhile, for the same input, when asked about the underlying paper (objects which reflected the color for the input), the responses showed color constancy. We believe that our request to report the target's presence or absence is more like the request to report on the paper object, thus probing inference.

It is in principle possible that the bias in the observers' reports did not arise from the inference stage (which gives $P(yes|C_t)$, or more strictly, $P(yes|x_t)$), but from the subsequent decision stage, when a threshold value P_{th} is chosen such that a response "yes" or "no" is given if $P(yes|x_t) > P_{th}$ or otherwise respectively[43]. The decision bias would be manifested in the choice of P_{th}, e.g., $P_{th} = 0.5$, 0.1, or 0.9. Our experiments can not distinguish between these two types of biases. However, if the bias was indeed only in the decision (in terms of P_{th}), then the inference $P(yes|x_t)$ is independent of the context. Without any insight on how contexts bias the decision threshold P_{th}, the decision bias has to be modelled by introducing one model parameter for each contextual condition (defined by a particular combination of the configuration and contrast C_c of the context), in addition to the model parameters for the unbiased inference $P(yes|C_t)$ or $P(yes|x_t)$ shared by

all contextual conditions. Hence decision bias is a less parsimonious model to account for our data since it would require more model parameters than our model of inference bias. In addition, other than a numerical value P_{th}, the decision bias does not give any insight in why and how the decision should be biased by context when the inference is unbiased. It is most likely that our measured yes rate results from the combined effect of (1) a context specific inference bias in the posterior $P(yes|x_t)$, and, (2) a context independent decision bias in P_{th} arising from observers' wish to give the "yes" response in roughly half of all trials. As our task can not distinguish between these two biases, our fitted values for $P(yes)$ manifest the combined effect from both biases, as discussed in the Results section.

One may wonder whether the sensitivities in the 2AFC task could be derived as the derivatives of the psychometric function (the yes rate) observed in our yes-no task using the same stimuli[44]. The answer is not so. First, it is likely, as discussed earlier, that different mechanisms are involved in input discrimination (for assessing sensitivity) and object inference, such that the input sensitivities and yes rates may not be so simply related. The second reason for the negative answer is the following. The 2AFC tasks were typically performed in blocked sessions, each having only a single contextual condition, while our yes-no design randomly interleaves trials of the different contextual conditions, such that observers compensate fewer "yes" responses in one contextual condition by more "yes" responses in another within a single session. Hence, the yes rates in one context is influenced by the other contexts interleaved within the same experimental session. Consequently, the three yes rate curves in the same no context condition in our three experiments are different from each other, and none of them could be simply related to the sensitivites in the 2AFC task performed in blocked trials. Recently, Polat and Sagi[45] also found, by a yes-no design, different biases to respond "yes" for a gabor target in different colinear contexts (in terms of different target-context distances), when trials of different contextual conditions were interleaved. In comparison with their study, the current study additionally reveals how this bias depends on the contextual contrast, how a Bayesian model can explain the data, and our additional data and the model have enabled us to show that there is no colinear facilitation or suppression of target contrast in such a visual inference task.

In our model, the parameters k and σ_n reflect the brain's internal model of the sensory world and its encoding. This internal model adapts quickly to the statistics of the external inputs[46], in particular, to the collection of the inputs presented in an experiment. Therefore, our different experiments, using different collections of stimuli, will evoke different internal models, as manifested by the different values of the model parameters k and σ_n.

Our observers seemed unconsciously to use prior beliefs induced by context, despite our instructions informing them that the context was irrelevant to the task. Furthermore, they could quickly switch from one prior to another as the context changes from one trial to another. However, these different priors are only different from the perspective of the target alone. When combining target and context as a whole, the joint prior probability of the visual input in principle arises from the same underlying probability distribution[47] of visual inputs derived from the ecological experience of the observers. Combining computational modeling with psychophysical experiments using easily controlled stimuli, the method in this study enables linking the visual inference behav-

ior with plausible neural substrates. The current study is only a beginning of using such a method, which can be a powerful tool in future studies of visual inference processes.

4 Materials and Methods

4.1 Stimuli

The stimuli were shown on a gamma-corrected 21 inch Sony GDM-F520 monitor using 14-bits luminance resolution. The viewing distance was 67.6 cm, and the screen width was 40 centimeters. All stimulus (target or contextual) bars were rectangular shapes of $0.9^o \times$ 0.165^o in visual angle, with a luminance L_{max} no smaller than the background luminance of $L_{min} = 15.6cd/m^2$ such that the contrast of a bar is $(L_{max} - L_{min})/(L_{max} + L_{min})$; The vertical target bar was always at the display center. Pilot experments established that the contrast detection threshold without contexts is around $C_t = 0.005$, measured in a 2AFC task with the stair case method. The stimuli were always presented with four black discs, of size 0.2^o in diameter, at the four corners of an imaginary square centered at the target location, the side of this square is 1^o in visual angle. These four black discs alone on the background also served as the fixation stimulus.

4.2 Procedure

Each observer was between 18-40 years old, had normal or corrected-to-normal vision, and participated in only one experment. The experiments were carried out in a dimly lit room. Each trial began with the fixation display for 500 ms, followed by the test stimulus display for 80 ms together with an auditory beep, which is then followed by the fixation display which stayed on waiting for observers' button press response to indicate whether they perceived the target or not in the trial. No feedbacks were given regarding whether their responses were correct. The next trial started 800 ms after the button press. Twenty randomly selected trials were performed before data collection for each observer before each session. Each experimental session randomly interleaved different stimulus conditions, such that the observers could not predict beyond chance the target contrast C_t, nor the contextual configuration and contrast C_c before each trial.

5 Acknowledgement

We thank Gatsby Charitable Foundation and a Cognitive Science Foresight grant BBSRC #GR/E002536/01 for funding, Joshua A Solomon and Mike Morgan for very helpful discussions and help on references, the two anonymous reviewers for their constructive comments, and Peter Dayan for reading the manuscript and comments.

6 Appendix

6.1 Formulation of the Bayesian influence and decision

Here we formulate our Bayesian inference and decision model in more detail. In a single trial, x_t and x_c are the neural responses to the target and the context respectively. The target stimuli is uniquely described by the target contrast C_t, as its other aspects (orientation, location, etc) are fixed. The contextual input is determined by both its contrast C_c and its spatial configuration S_c (describing orientation & location). Neural and input noise make x_t a random variable according to a conditional probability $P(x_t|C_t)$ of x_t given C_t, and similarly, x_c according to $P(x_c|C_c, S_c)$. The brain infers whether x_t is caused by a target or noise for the observer to respond "yes" or "no" to the question "is the target present?". This inference is partly based on the brain's internal model, expressed in conditional probability, $P(x_t|yes)$ or $P(x_t|no)$, of how likely x_t can be by target or non-target cause, when the brain assumes the target is present or abstract respectively. Contextual influences on the internal model $P(x_t|yes)$ is indicated by adding a subscript x_c, in $P_{x_c}(x_t|yes)$, denoting that $P(x_t|yes)$ is parameterized by x_c (we assume for simplicity that the context does not influence $P(x_t|no)$). The inference is also partly based on the context dependent prior probability $P_{x_c}(yes)$, assumed by the brain, that a target bar should be present. By the Bayesian formula, the brain infers from x_t that the probability for a target to be present in this trial is

$$P(yes|x_t) = \frac{P_{x_c}(x_t|yes)P_{x_c}(yes)}{P_{x_c}(x_t|yes)P_{x_c}(yes) + P(x_t|no)(1 - P_{x_c}(yes))} \tag{7}$$

If the observer responds "yes" or "no", the probability of error is $1 - P(yes|x_t)$ or $P(yes|x_t)$ respectively. To minimize error (assuming that the error rate is the loss function for the decision), the optimal response is "yes" when $P(yes|x_t) > 0.5$ and "no" otherwise. Averaging over many trials of fluctuating neural and observer responses, we obtain the probability of "yes" response for a given target and contextual stimuli (C_t, C_c, S_c).

$$P(yes|C_t) = \int dx_t P(x_t|C_t) \int dx_c P(x_c|C_c, S_c) H(P(yes|x_t) - 0.5) \tag{8}$$

where $H(.)$ is a step function such that $H(x) = 1$ or 0 when $x > 0$ or otherwise respectively.

The posterior probability $P(yes|C_t)$ should depend on C_t, C_c, and S_c, with some functional parameters derived from the functional parameters in $P_{x_c}(x_t|yes)$, $P(x_t|no)$, $P(x_t|C_t)$, $P(x_c|C_c, S_c)$, and $P_{x_c}(yes)$. For our purpose, all we need is to parameterize the dependence of $P(yes|C_t)$ on C_t, C_c, and S_c by a suitable phenomenological model that has enough parameters, but, applying Occam's razor, not too many. Hence, we use the following Ansatz

$$P(yes|C_t) = \frac{P(C_t|yes)P(yes)}{P(C_t|yes)P(yes) + P(C_t|no)(1 - P(yes))} \tag{9}$$

using three phenomenological parameters: one is $P(yes)$ to parameterize the dependence on S_c, and the other two σ_n and k, parameterizing the dependence on C_c and C_t, are defined in the definition of $P(C_t|yes)$ and $P(C_t|no)$ as

$$P(C_t|no) = \frac{\exp(-C_t/\sigma_n)}{N_n}, \tag{10}$$

$$P(C_t|yes) = \frac{\exp(-|C_t - C_c|/(k \cdot C_c))}{N_y}, \tag{11}$$

where N_n and N_y are normalization constants such that $\int_0^1 dC_t P(C_t|no) = 1$ and $\int_0^1 dC_t P(C_t|yes) = 1$

While an Ansatz is typically justified by its suitability in accounting for the data, as demonstrated in the main text for the Ansatz above, here we provide some motivations behind this Ansatz. For ease of presentation, we abbreviate the integration $\int dx_t P(x_t|C_t) \int dx_c P(x_c|C_c, S_c)$ over internal variables by $\oint d\mathcal{X}$. Equation (8) suggests an approximation $P(yes|C_t) \approx \oint d\mathcal{X} P(yes|x_t)$. Then, the certainty equivalent approximation of this equation suggests equation (9), with approximations $P(C_t|yes) \approx \oint d\mathcal{X} P_{x_c}(x_t|yes)$, $P(C_t|no) \approx \oint d\mathcal{X} P(x_t|no)$, and $P(yes) \approx \oint d\mathcal{X} P_{x_c}(yes)$. These approximations do not need to be accurate, since the model parameters are to be fitted by behavioral data rather than derived from integrating these equations. They simply serve to suggest that equation (9) is a suitable phenomenological model, with $P(yes)$ the phenomenological prior, and $P(C_t|yes)$ or $P(C_t|no)$ the phenomenological conditional probability, assumed by the brain, that the input contrast should be C_t for a target bar or otherwise, respectively.

The model $P(C_t|no) \propto \exp(-C_t/\sigma_n)$ is motivated by the brain's internal model that, without a target, the perceived C_t is more likely zero than another value $C_t > 0$. Under a simplifying assumption that $P_{x_c}(yes)$ is influenced only by the contextual configuration S_c, $P(yes) \approx \oint d\mathcal{X} P_{x_c}(yes)$ becomes a mere parameter for each contextual configuration. Meanwhile, the form of $P(C_t|yes)$ is motivated by its approximation $\int dx_t dx_c P(x_t|C_t) P(x_c|C_c, S_c) P_{x_c}(x_t|yes)$ as follows. Physiologically[48, 49], the encoding neural response is roughly a sigmoid-like function of the logarithm of input contrast, i.e., $x_c = g(\log C_c)+$ noise, with $g(.)$ denoting this sigmoid like function. Thus, $P(x_c|C_c, S_c)$ peaks around $x_c = g(\log C_c)$ and decreases with $|x_c - g(\log C_c)|$ (this is presumably the basis of the Weber law: that the behavorially just discriminable contrast difference between a pedestal contrast and a second contrast is proportional to the pedestal contrast). Similarly, $P(x_t|C_t)$ peaks around $x_t = g(\log C_t)$ and descreases with $|x_t - g(\log C_t)|$. Assuming again for simplicity that $P_{x_c}(x_t|yes)$ is only influenced by the contextual contrast C_c, the response x_c to a context bar makes the brain expect that x_t should resemble x_c (which are after all examples of neural responses to stimulus bars), making $P_{x_c}(x_t|yes)$ peak around $x_t \approx x_c$. Combining these observations, $P(C_t|yes) \approx \int dx_t dx_c P(x_t|C_t) P(x_c|C_c, S_c) P_{x_c}(x_t|yes)$ as a function of C_t and C_c should depend approximately on the difference $\log C_c - \log C_t$ or the ratio C_t/C_c. The model $P(C_t|yes) \propto \exp(-|C_t - C_c|/(k \cdot C_c))$ suits such a form, whereas an alternative like $P(C_t|yes) \propto \exp(-|C_t - C_c|/\sigma_c)$ (with a fixed parameter σ_c) would not.

Other additional variabilities, such as the perceived locations of the stimulus, would behave analogously to the internal variables x_t and x_c which should be integrated over, as in equation (8), to arrive at the experimental observation $P(yes|C_t)$. One could generalize the definition of x_t and x_c, making each a vector with multiple components for multiple variables, e.g., the first component of x_t for the neural response to the target contrast, the second the neural representation for the target location, etc. Repeating the above derivations would lead us again to equation (9). By not detailing these additional variables, we are assuming that they will not significantly affect the suitability of our phenomenological model in equations (9- 11). The fitted model parameters

manifest the combined effects from all the variables x_t and x_c, even though only a fraction of them play a dominant role.

6.2 Considering contextual influences on the encoding process

Context could affect the target encoding by changing $P(x_t|C_t)$. We consider a situation when context could change input sensitivity such that the encoding neurons respond as if the input contrast is effectively $C_t^{effective} = C_t + \Delta C_t \neq C_t$. If $P(x_t|C_t)$ without the context takes a functional form $P(x_t|C_t) = F(x_t, C_t)$ where $F(.)$ is some function of x_t and C_t, the contextual influence makes $P(x_t|C_t) = F(x_t, C_t^{effective})$. This motivates the phenomenological formulation to modify the right hand side of equation (9) such that C_t is replaced by $C_t^{effective}$. This contextual influence in encoding can then be phenomenologically modelled by parameterizing the dependence of ΔC_t on the context as, e.g., $\Delta C_t \approx \gamma C_c$, as done in the main text.

6.3 Proof of $P_{C_{c2}}(C_t|yes) > P_{C_{c1}}(C_t|yes)$ when $C_t < C_{c2} < C_{c1}$ for contextual contrasts C_{c1} and C_{c2} concerned

We use subscript C_c in $P_{C_c}(C_t|yes)$ to denote that this probability of target contrast C_t is parameterized by contextual contrast C_c. When $P_{C_c}(C_t|yes) = \exp(-|C_t - C_c|/(kC_c))/N_y$ with $N_y = kC_c[2 - \exp(-1/k) - \exp(-1/(kC_c) + 1/k)]$, we have, denoting N_y for C_{c1} and C_{c2} as $N_y(1)$ and $N_y(2)$ respectively,

$$\frac{P_{C_{c2}}(C_t|yes)}{P_{C_{c1}}(C_t|yes)} = \exp[(C_t/k)(\frac{1}{C_{c2}} - \frac{1}{C_{c1}})]\frac{N_y(1)}{N_y(2)} \tag{12}$$

since $\exp[(C_t/k)(\frac{1}{C_{c2}} - \frac{1}{C_{c1}})] \geq 1$, $\frac{P_{C_{c2}}(C_t|yes)}{P_{C_{c1}}(C_t|yes)} > 1$ if $\frac{N_y(1)}{N_y(2)} > 1$. Note that $N_y(i) \equiv \psi_i + \phi_i$ where $\psi_i = \int_0^{C_{c1}+C_{c2}} \exp[-|C - C_{ci}|/(kC_{ci})]dC$ and $\phi_i = \int_{C_{c1}+C_{c2}}^1 \exp[-(C - C_{ci})/(kC_{ci})]dC$ for $i = 1, 2$. Changing integration variable $C \to C_{c1} + C_{c2} - C$ in ψ_1 we have $\psi_1 = \int_0^{C_{c1}+C_{c2}} \exp[-|C - C_{c2}|/(kC_{c1})]dC$, hence $\psi_1 > \psi_2 = \int_0^{C_{c1}+C_{c2}} \exp[-|C - C_{c2}|/(kC_{c2})]dC$ given $C_{c1} > C_{c2}$. Meanwhile, $\exp[-(C - C_{c1})/(kC_{c1})] > \exp[-(C - C_{c2})/(kC_{c2})]$ for all $C \geq C_{c1} + C_{c2}$. Hence $\phi_1 > \phi_2$ as long as $C_{c1} + C_{c2} < 1$, i.e., the contextual contrasts are not super-saturating. This applies to all of our experimentally used contrasts $C_c \leq 0.4$. (In fact, when contextual contrasts are beyond this range, neural responses are saturating and our phenomenological model of the form $P_{C_c}(C_t|yes) \propto \exp(-|C_t - C_c|/(kC_c))$ may or may not be the most suitable). Hence $N_y(1) > N_y(2)$, and then $P_{C_{c2}}(C_t|yes) > P_{C_{c1}}(C_t|yes)$.

References

[1] Zhou H, Friedman HS, von der Heydt R. (2000) Coding of border ownership in monkey visual cortex. *J Neurosci.* 20(17):6594-611.

[2] Zhaoping L. (2005) Border ownership from intracortical interactions in visual area v2. *Neuron* 47(1):143-53.

[3] Henderson JM. and Hollingworth A. (1999) High-level scene perception. Ann. Rev. Psycho. 50, 243-271.

[4] Brown RO and MacLeod DI (1997) Color appearance depends on the variance of surround colors *Current Biology* 7(11):844-849.

[5] Nelson JI, Frost BJ. (1985) Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex. *Exp Brain Res.* 61(1):54-61.

[6] Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. Neuron, 15, 843-856.

[7] Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Noricia, A. M. (1998). Collinear stimuli regulate visual responses depending on cells contrast threshold. Nature, 391, 580-584.

[8] Polat & Sagi, 1993 Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Res.* 33(7):993-9.

[9] Morgan, M. J., & Dresp, B. (1995). Contrast detection facilitation by spatially separated targets and inducers. Vision Research, 35, 1019-1024.

[10] Wehrhahn, C., & Dresp, B. (1998). Detection facilitation by collinear stimuli in humans: dependence on strength and sign of contrast. Vision Research, 38, 423-428.

[11] Snowden, R. J., & Hammett, S. T. (1998). The effects of surround contrast on contrast threshoulds, perceived contrast and contrast discrimination. Vision Research, 38, 1935-1945.

[12] Yu, C., &, Levi, D. M. (2000). Surround modulation in human vision unmasked by masking experiments. Nature Neuroscience, 3, 724-728.

[13] Huang P.C., Hess, R.F., & Dakin S. C. (2006) Flank facilitation and contour integration: difference sites. *Vision Research* 46:3699-3706.

[14] Kersten D., Mamassian P., and Yuille A. (2004) Object perception as bayesian inference. *Annual Review of Psychology* 55:271-304.

[15] Torralba A. and Sinha P. (2001) Statistical context priming for object detection. *International conference on Computer Vision* http://doi.ieeecomputersociety.org/10.1109/ICCV.2001.10051

[16] Sinha, P. & Poggio, T. (1996). I think I know that face..., *Nature* 384(6608):404.

[17] Golz J, MacLeod DI. (2002) Influence of scene statistics on colour constancy. *Nature* 415(6872):637-40.

[18] Macmillan, N. A., & Creelman, C. D. (2005). Detection theory: A users guide (2nd Edition).Lawrence Erlbaum Associates, Inc., US.

[19] Dayan P and Abbott L.F. (2001) *Theoretical Neuroscience*, MIT press.

[20] Shapley R and Reid RC (1985) Contrast and assimilation in the perception of brightness. *Proc. National Acad. of Sci. USA* 82(17):5983-5986.

[21] van Lier R. and Wagemans J. (1997) Perceptual grouping measured by color assimilation: regularity versus proximity. *Acta Psychologica* 97:37-70.

[22] Singer B. and D'Zmura M. (1993) Color contrast induction. *Vision Res.* 34(23):3111-3126.

[23] Solomon JA, Felisberti FM, and Morgan MJ. (2004) Crowding and the tilt illusion: Toward a unified account. *Journal of Vision* 4, 500-508.

[24] Murakami I and Shimojo S. (1993) Motion capture changes to induced motion at higher luminance contrasts, smaller eccentricities, and larger inducer sizes. *Vision Res.* 33(15):2091-2107.

[25] Barlow HB, 1961, "Possible principles underlying the transformations of sensory messages." In: Sensory Communication W.A. Rosenblith, ed., Cambridge MA, MIT Press, pp. 217-234.

[26] Zhaoping L. (2006) Theoretical Understanding of the early visual processes by data compression and data selection. *Network: Computation in neural systems* 17(4):301-334.

[27] Biederman I, Mezzanotte RJ, Rabinowitz JC. (1982) Scene perception: detecting and judging objects undergoing relational violations. *Cognit Psychol.* 14(2):143-77

[28] Chun MM. (2000) Contextual cueing of visual attention. *Trends Cogn. Sci.* 4(5):170-178.

[29] Albright TD and Stoner GR. (2002) Contextual influences on visual processing *Annu. Rev. Neurosci.* 25:339-79.

[30] Nakayama K. He Z. and Shimojo S. (1995) Visual surface representation: a critical link between lower-level and higher-level vision. In Kosslyn, S. and Osherson D., editors, Visual cognition: An invitation to cognitive science, vol 2. (2nd Edition), pages. 1-70. MIT press, Cambridge MA, USA.

[31] Allman J, Miezin F, McGuinness E. (1985) Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev. Neurosci.* 8:407-30.

[32] Hess RF, Dakin SC, Field DJ. (1998) The role of "contrast enhancement" in the detection and appearance of visual contours. *Vision Research* 38(6):783-7.

[33] Li W, Piech V, Gilbert CD. (2004) Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.* 7(6):651-7.

[34] Li W, Piech V, Gilbert CD. (2006) Contour saliency in primary visual cortex. *Neuron* 50(6):951-62.

[35] von der Heydt R, Peterhans E, Baumgartner G. (1984) Illusory contours and cortical neuron responses. *Science* 224(4654):1260-2.

[36] Bakin JS, Nakayama K, Gilbert CD (2000) Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *J. Neuroscience* 20(21):8188-8198.

[37] Zhaoping L. (2002) Pre-attentive segmentation and correspondence in stereo. *Philos Trans R Soc Lond B Biol Sci* 357(1428):1877-83.

[38] Roe AW, Lu HD, and Hung CP (2005) Cortical processing of a brightness illusion. *Proc. Natl. Acad. Sci. USA* 102:3869-74.

[39] Sasaki Y and Watanabe T (2004) The primary visual cortex fills in color. *Proc. Natl. Acad. Sci. USA* 101:18251-56.

[40] Boyaci H, Fang F, Murray SO, and Kersten D. (2007) Responses to lightness variations in early human visual cortex. *Current Biology* 17:989-993.

[41] Hills JM, and Brainard DH (2007) Distinct mechanisms mediate visual detection and identification *Current Biology* 17(1714-1719).

[42] Arend L. and Reeves A. (1986) Simultaneous color constancy. *J. Opt. Soc. Am. A* 3(10):1743-51.

[43] Pelli DG. (1985) Uncertainty explains many aspects of visual contrast detection and discrimination. *J. Opt. Soc. Am. A* 2(9):1508-32.

[44] Solomon, J. A., Watson, A. B., and Morgan, M. J. (1999). Transducer model produces facilitation from opposite-sign flankers. Vision Research, 39, 987-992.

[45] Polat, U, & Sagi, D. (2007) The relationship between the subjective and objective aspects of visual filling-in. *Vision Research* 47(18):2473-81.

[46] Schwartz O, Hsu A, Dayan P. (2007) Space and time in visual context. *Nature Review Neurosci.* 8(7):522-35.

[47] Yu, AJ, Dayan, P & Cohen JD. (2007) Bayesian models of dynamic attentional selection. Presented at COSYNE conference, 2007. available at http://cosyne.org/c/images/c/ce/Cosyne-poster-I-2.pdf

[48] Albrecht DG, Hamilton DB. (1982) Striate cortex of monkey and cat: contrast response function. *J Neurophysiol.* 48(1):217-37.

[49] Valeton M.J. and van Norren D. (1983) Light adaptation of primate cones: an analysis based on extracellular data. *Vision Research* 23, 1539-47.