# Different processing strategies underlie voluntary averaging in low and high noise

**Rémy Allard** Laboratoire Psychologie de la Perception, Paris, France

**Patrick Cavanagh** Laboratoire Psychologie de la Perception, Paris, France

An underlying assumption of the external noise paradigm is that the same processing strategy operates whether the dominating noise source comes from the observer (i.e., internal) or the stimulus (i.e., external). Here, we challenged this noise-invariant processing assumption for a particular variant of the external noise paradigm—the voluntary averaging paradigm—where processing is characterized by the efficiency of averaging across the samples. The task consisted in discriminating the mean orientation of four distinctly perceived Gabors, and the external noise corresponded to orientation-jitter added to these Gabors. The averaging efficiencies for the four-sample case were measured by comparing discrimination thresholds of the average orientation of four Gabors to a baseline with a single Gabor. In high noise, orientation discrimination thresholds were *better* when 4 Gabors rather than 1 were presented, showing efficient averaging. But in absence of external noise, presenting four identically oriented Gabors rather than one did not improve performance, showing that subjects no longer averaged the individual estimates. We conclude that the averaging process operating in high noise does not operate in low noise because there is no reason to voluntarily average stimuli that appear identical. This conclusion implies that the processing strategy can change depending on the external noise level, an implication which violates the noise-invariant processing assumption underlying the averaging paradigm.

Keywords: external noise, orientation discrimination, averaging

Citation: Allard, R., & Cavanagh, P. (2012). Different processing strategies underlie voluntary averaging in low and high noise. *Journal of Vision, 12*(11):6, 1–12, http://www.journalofvision.org/content/12/11/6, doi:10.1167/12.11.6.

## Introduction

External noise paradigms are often used to investigate visual processing (e.g., Allard & Faubert, 2006, 2008; Bennett, Sekuler, & Ozin, 1999; Dosher & Lu, 2004; Lu & Dosher, 2004; Lu & Dosher, 2008; Pardhan, 2004; Pardhan, Gilchrist, & Beh, 1993; Pelli, 1981, 1990; Pelli & Farell, 1999). The paradigm characterizes processing stages by analyzing the relative effects of observer (internal) and the stimulus (external) noise. An underlying assumption of the external noise paradigm is that the processing strategy does not depend on the noise level. This noise-invariance is generally taken for granted, but we have recently shown that the processing strategy can change drastically depending on the type of external noise for a detection task in luminance noise (Allard & Cavanagh, 2011). Here we show that the violation of the noise-invariant processing assumption was not a peculiarity of contrast detection in luminance noise as we observed a processing strategy shift for another version of the external noise paradigm—voluntary averaging.

In voluntary averaging paradigms (Dakin, 1999, 2001; Dakin, Bex, Cass, & Watt, 2009; Dakin, Mareschal, & Bex, 2005; Solomon, 2010) observers estimate the mean of a Gaussian probability density function (pdf) based on a given number of samples drawn from it. The standard deviation (*SD*) of the pdf represents the external noise independently added to each sample. In the current study, the task consisted in estimating whether the mean orientation of a set of Gabors was tilted clockwise or counterclockwise from vertical (Figure 1b). The processing in this case is analyzed with an averaging model (Figure 1b), which in this example is a linear amplifier model (Figure 1a) adapted to averaging paradigms. The averaging model has two processing stages: (a) sample estimation (its precision is modeled as internal noise) and (b) the averaging of these estimates. When the *SD* of the external pdf is large (i.e.., high external noise), the precision of the sample estimates has negligible impact (external noise dominates internal noise) and performance only depends on the averaging efficiency. But when the *SD* of the pdf is small (i.e., low external noise), the performance is assumed to depend on both the precision of the sample estimates (observer's internal noise assumed to be uncorrelated across samples) and the averaging of these estimates. Thus, the averaging efficiency is derived from the performance in high noise while the precision of sample

doi: 10.1167/12.11.6 Received December 22, 2011; published October 9, 2012 ISSN 1534-7362 

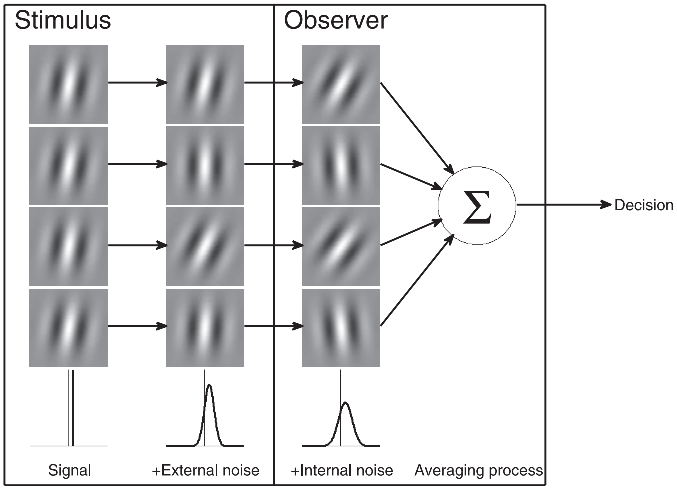


Figure 1. a) The linear amplifier model. A stimulus is composed of a signal to which external noise is added. The visual system has its own internal noise, which is modeled as additive internal noise. The calculation efficiency reflects the signal-to-noise ratio required to perform the task. b) The averaging model (i.e., the linear amplifier model adapted to an averaging paradigm) for an orientation-averaging task. All samples are composed of a common orientation (the signal) to which an independent orientation imprecision is added (external noise). The observer is assumed to estimate the orientation of each sample more or less precisely (internal noise added to each sample) and then suboptimally averages these estimates (averaging efficiency).

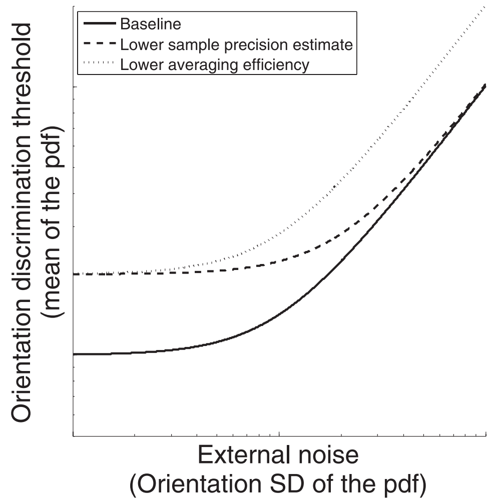


Figure 2. Predictions based on the averaging model. More internal noise (i.e., lower precision estimate of each sample) would affect performance when the dominating noise source is internal (i.e., low external noise) but have negligible effect when it is external (i.e., high external noise). A lower averaging efficiency would affect performance independently of the dominating noise source, i.e., at all noise levels.

estimates can be calculated from the performance in low noise if it assumed that the averaging efficiency is the same in low and high noise. Thus, based on the averaging model, both the precision of the sample estimates (internal noise) and averaging efficiency can be measured by evaluating the discrimination thresholds when the dominating noise source comes from the observer (i.e., internal) and the stimulus (i.e., external).

Such an averaging paradigm (which is a form of the external noise paradigm) is typically used to identify whether a given manipulation affects the precision of the sample estimates or the averaging efficiency. Based on the averaging model, a manipulation affecting sample precision would affect performance when the dominating noise source is internal (i.e., low external noise) but not when it is external. On the other hand, a manipulation affecting the averaging efficiency would affect performance independently of the dominating noise source (internal or external), i.e., performance would be affected in both low and high external noise (Figure 2). For instance, Dakin and colleagues (2009) found, for an orientation-averaging task, that crowding affected performance only in low noise and that attentional load affected performance in both low and high noise. Based on the averaging model, they concluded that crowding and attentional load have dissociable effects: crowding affects the precision of local estimates whereas attentional load affects averaging efficiency.

It is important to note that an attribute (e.g., orientation) can be averaged either voluntarily or involuntarily (see Dakin et al., 2009). Involuntary averaging occurs when the samples are unavoidably pooled, for example, under crowding conditions when the samples are near one another in periphery and are difficult to report independently. Voluntary averaging occurs when the observer averages elements that are more widely spaced so that they could be reported independently. In the current study, each sample (Gabor) was spaced far enough apart to avoid crowding, so the averaging process was voluntary. This distinction is important because the current study addresses the nature of the *voluntary* averaging process that is engaged when all samples appear identical as opposed to when they appear to be different.

The noise-invariant processing assumption implies that the same voluntary averaging process operates in both low and high noise, i.e., whether the dominating noise source is internal or external. In high noise, the internal noise has a negligible impact and observers average samples that differ due to external noise; and, vice versa, in low noise, the external noise has a negligible impact and observers average samples that differ, if at all, only due to internal noise. Thus, the voluntary averaging paradigm assumes that we perceive identical samples as slightly different due to noisy sample estimates (i.e., internal noise) and therefore we must average them to find a group estimate.

However, voluntarily averaging different samples is a demanding task, and we can easily imagine that observers might see no reason to average samples that appear to be identical. Indeed, we do not often perceive identically oriented Gabors as having slightly different orientations except when their signal strength is weak (with brief or noisy presentation). Indeed, we even perceive slightly different orientations as being identical (Morgan, Chubb, & Solomon, 2008). This could be due to a thresholding mechanism preventing us from perceiving our internal noise (Morgan, et al., 2008; Ross & Burr, 2008) or to a grouping phenomenon.

If we do not take the effort to voluntarily average samples when they appear identical, then there are clearly different processing strategies operating when the samples appear identical (i.e., in low noise) compared to when they appear different (i.e., in high noise). Such a strategy shift would violate the noise-invariant processing assumption underlying the voluntary-averaging paradigm. Specifically, we can no longer assume that observers have the same averaging efficiency in both low and high noise (as required by the averaging model).

The present study compared the averaging efficiencies in a voluntary averaging task when the dominating noise source was internal and external. Based on the averaging model, the averaging efficiency is usually measured only in high noise (i.e., when the precision of the sample estimates has negligible impact and performance only depends on the averaging efficiency). This averaging efficiency is assumed to be the same in low noise as it is in high noise. In the current study, we directly measured averaging efficiencies in both low and high noise. To anticipate our results, we found that observers could efficiently average many samples in high noise, but showed no evidence of averaging (same performance given one and four samples) in low noise. We conclude that the averaging process operating in high noise does not operate in low noise, violating the noise-invariant processing assumption underlying the external noise paradigm.

## Averaging efficiency

In high noise, the precision of the sample estimates has negligible impact as the external noise dominates internal noise ($N_{ext} \gg N_{int}$) and so performance only depends on the averaging efficiency. Thus, performance in high noise is usually used to measure the averaging efficiency. Typically, it is quantified as the number of samples an ideal observer would need to have the same performance as the observer. The equality between the ideal observer discrimination threshold given $n$ samples at the noise level $N_{ext}$ ($I_n(N_{ext})$) and the observer's discrimination threshold given $m$ samples at the same noise level ($O_m(N_{ext})$) can be represented as:

$$O_m(N_{ext}) = I_n(N_{ext}). \quad (1)$$

Since the ideal discrimination threshold is proportional to the square root of the number of samples, then

$$I_m(N_{ext}) = \frac{I_n(N_{ext})}{\sqrt{n}}. \quad (2)$$

Thus,

$$O_m(N_{ext}) = \frac{I_n(N_{ext})}{\sqrt{n}} \quad (3)$$

and the number of samples ($n$) an ideal observer would need to have the same discrimination threshold as the observer is:

$$n = \left(\frac{I_1(N_{ext})}{O_m(N_{ext})}\right)^2. \quad (4)$$

This function is typically used to quantify the averaging efficiency in high noise and, based on the averaging model (Figure 2), this efficiency is assumed to be the same in low noise.

The goal of the present study was to challenge the averaging model by measuring the averaging efficiencies in both low and high noise. To do so, we compared the observer's performances given 1 and $m$ samples. Given that the performance for an averaging task in high noise only depends on the averaging efficiency and averaging one sample is trivial, the performance of an observer should be just as good as that of an ideal observer when only one sample is presented (which was confirmed for most of our observers). Thus, we expect

$$O_1(N_{ext}) \approx I_1(N_{ext}). \quad (5)$$

As a result, the averaging efficiency given $m$ samples can be measured relative to the performance given one sample:

$$n \approx \left(\frac{O_1(N_{ext})}{O_m(N_{ext})}\right)^2. \quad (6)$$

If we assume that the precision of the sample estimates is independent of the set size (which is a reasonable assumption; see Solomon, 2010), then this formulation can also be used to measure the averaging efficiency in low noise. Indeed, the performance ratio given one sample relative to $m$ samples will only depend on the factors that are set-size dependent, i.e., the averaging efficiency. As a result, Equation 6 can be used to measure the averaging efficiencies in both low and high noise.

## Experiment 1: Averaging efficiency and the averaging model

The goal of the first experiment was to examine the assumptions of the averaging model (Figure 1) for a mean orientation discrimination task. This model assumes that the same averaging efficiency in low ($N_{ext} \ll N_{int}$) and high ($N_{ext} \gg N_{int}$) noise, i.e., whether the main orientation imprecision comes from the observer or the stimulus, respectively. We therefore measured the averaging efficiency in both low and high noise by comparing discrimination thresholds given four samples ($O_4(N_{ext})$) with discrimination threshold given one sample ($O_1(N_{ext})$).

### Method

#### Observers

Three naïve observers and an author participated in this experiment. They all had normal or corrected-to-normal vision.

#### Stimuli

The stimuli were composed of either one or four Gabors, which were displayed 8° of visual angle to the left, right, above, and below fixation (Figure 3). When only one Gabor was presented, its position was randomly assigned to one of these four positions so that observers could not anticipate its position. The Gabors were presented for 200 ms, which was too brief for the observer to saccade to the target (Hallett, 1986). The frequency of the Gabors was 2 cpd; their spatial envelope was a Gaussian with a *SD* of 0.33 degrees of visual angle (dva); their contrast was maximized and their phases were randomized. The orientations of the Gabors were selected from a Gaussian distribution. The mean of the distribution depended on the staircase procedure to measure orientation discrimination thresholds (see Procedure section below). The *SD* of the distribution was either 0° or 16°, i.e., no or high external noise.

#### Apparatus

Stimuli were computer generated and presented on a gamma-linearized 22” Formac ProNitron 22,800 CRT monitor with a mean luminance of 42 cd/m$^2$ and a refresh rate of 85 Hz. The observer’s head was supported by a chin rest positioned 65 cm from the display. The monitor was the only light source in the room.

#### Procedure

Orientation discrimination thresholds were measured using a 2-down-1-up staircase procedure (Levitt, 1971). For each noise level ($SD = 0°$ and 16°) and each set size (1 and 4), three staircases were performed. Each staircase was ended after 12 inversions. The four staircases were performed three times in a pseudo-

Figure 3. Stimuli examples when four Gabors were presented in absence of external noise (left, i.e., all Gabors have the same orientation) or in high noise (right). Observers were asked to judge if the mean orientation was tilted clockwise or counterclockwise from vertical.

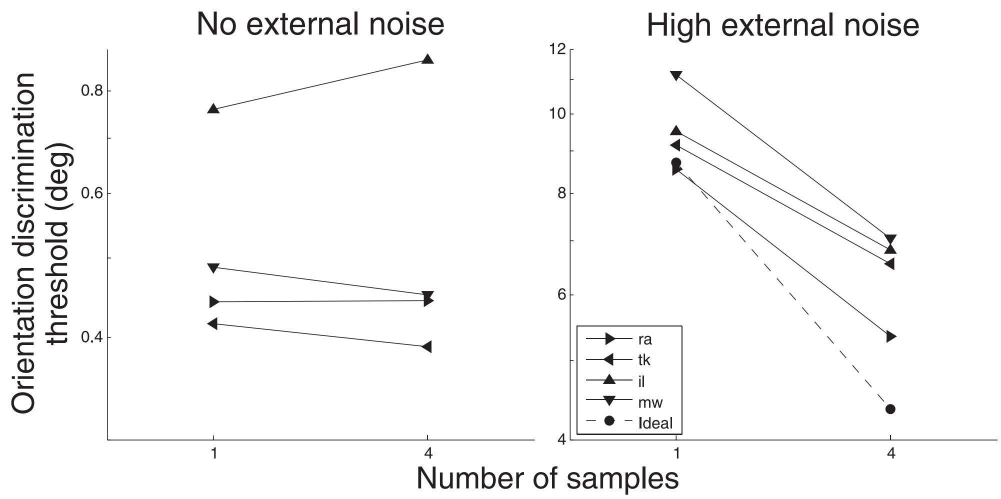


Figure 4. Orientation discrimination threshold measured in absence of external noise (left graph, $SD = 0°$) and in high external noise (right graph, $SD = 16°$) for four observers. The dotted line in high external noise shows the performance of an ideal observer ($I_1$ (16) and $I_4$ (16); see text for details).

random order. The orientation discrimination threshold was estimated by averaging the mean orientation of the last six inversions (step size of 0.05 log) of the three staircases. A feedback sound indicated the correctness of the response.

## Results and discussion

Figure 4 shows the orientation discrimination thresholds in absence of external noise (left) and in high noise (right) given one and four samples. Figure 5 shows the averaging efficiencies in absence of external noise and in high noise calculated from these thresholds (Equation 6). In high noise, observers were better at discrimination of the mean of the prior pdf given four samples compared to one sample (negative slope in the right graph of Figure 4 and an averaging efficiency above 1 in Figure 5). The averaging efficiency in high noise was about 2.2, meaning that observers were averaging across, on average, at least 2.2 samples. Indeed, these results could be explained by an ideal process averaging only 2.2 samples or by a suboptimal process averaging more samples. In absence of external noise, however, observers had similar thresholds whether one or four targets were presented (left graph of Figure 4 and an averaging efficiency near 1 in Figure 5). The different averaging efficiencies observed in absence of external noise and in high noise are incompatible with the averaging model presented above (Figure 1), which assumes that the averaging efficiency is independent of the external noise level.

Given one sample, an ideal observer estimates the mean of a pdf to be tilted clockwise or counterclockwise depending on whether the sample is tilted clockwise or counterclockwise, respectively. Given a *SD* of 16°, a Gaussian pdf must be 8.7° away from vertical so that 70.7% (the threshold criterion) of the samples will be tilted in the same direction as the mean of the pdf. Thus, the ideal discrimination threshold for an ideal observer given 1 sample and a *SD* of 16° ($I_1(16)$) is 8.7°. The ideal orientation discrimination threshold given four samples ($I_4(16)$) can be calculated using Equation 3.

Even though these results are inconsistent with the averaging model, they do not *necessarily* require that the processing strategy differed depending on the external noise level (which would be a violation of the

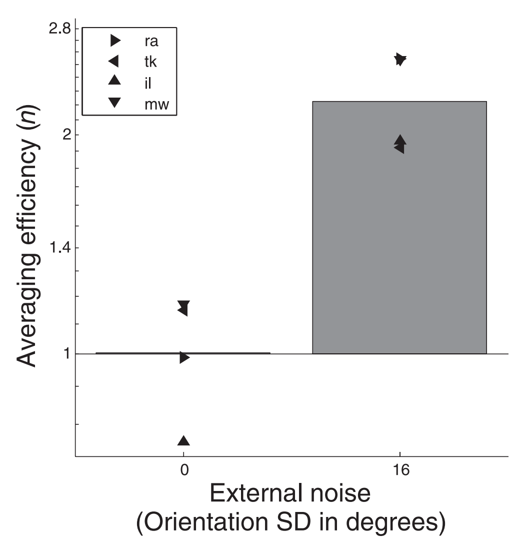


Figure 5. Averaging efficiencies (Equation 6) measured in absence of external noise ($SD = 0°$) and in high external noise ($SD = 16°$) for four observers. Note that the absence of averaging would result in an averaging efficiency of 1, i.e., the same orientation discrimination threshold given one or four samples.

noise-invariant processing assumption). Another model could explain both the substantial averaging efficiency in high noise and the absence of averaging efficiency in low noise without violating the noise-invariant processing assumption. The same averaging process could be operating in both low and high noise and substantially affect performance in high noise but not in low noise. The negligible impact of the averaging process in low noise could be due to correlated internal noise across samples or internal noise occurring after the averaging process for whatever reason (i.e., late noise; see Solomon, 2010). Both the correlation and the late noise effects could arise from a noisy internal vertical reference. In one case, if the processing strategy consisted in comparing each sample with a common vertical reference before averaging these comparisons, then a noisy internal vertical reference would result in early correlated noise across samples. In the other case, if the processing strategy consisted in comparing the averaged orientation across all samples with an internal vertical reference, then a noisy internal vertical reference would result in late noise (i.e., noise occurring after the averaging process). In either case, a noisy internal vertical reference could hide the effect of an averaging process without violating the noise-invariant processing assumption (the same processing strategy could be operating in both low and high noise). Thus, the results of the current experiment are incompatible with the simple averaging model presented above but do not necessarily violate the noise-invariant processing assumption.

Before addressing the question of correlated or late noise, we first conducted a control experiment to make sure that in absence of external noise, performance was limited by observers' performance and not by limits on the precision of the presented orientations imposed by the display itself. Indeed, if display resolution were the limiting factor setting the orientation discrimination thresholds, it would explain the similar orientation discrimination thresholds obtained with one and four samples. To test this, we used the same four Gabors presented in absence of noise, but instead of fixating at the display center, the observer (RA, one of the authors) looked directly at one of the Gabors. This change in gaze position improved his sensitivity substantially (40%), showing that orientation discrimination thresholds were limited by the observers' performance, not by display factors.

Note that when only 1 target was presented in high noise, orientation discrimination thresholds were close to the ideal performance (Equation 5) for three out of four observers (Figure 4). This shows that the averaging efficiency measured in high noise relative to the observers' performance when only one target was presented (Equation 6) is comparable to the conventional measure of the averaging efficiency relative to an ideal observer (Equation 4).

## Experiment 2: Controlling for orientation bias and observer's voluntary strategy

The main goal of the second experiment was to examine the noise-invariant assumption itself (not just the averaging model). To avoid the comparison with an internal vertical reference, each sample was composed of two nearby Gabors oriented in opposite orientations with the same magnitude (Figure 6) and the task consisted in determining whether the upper one was oriented clockwise or counterclockwise relative to the other. Estimating each sample by comparing two orientations minimizes both late and early correlated internal noise due to a noisy internal vertical reference. Indeed, given that observers estimate each sample by comparing two orientations, no comparison with an internal vertical reference is required after the averaging of sample estimates, and the impact of a common orientation bias to all Gabors should have negligible impact on orientation differences.

Another potential explanation of the absence of summation in low noise observed in the first experiment is that observers may have voluntarily changed their processing strategy depending on the known external noise levels (0° or 16°) since the conditions were blocked. Thus, observers may have based their decision on a single Gabor when they knew that all Gabors had the same orientation, and on many Gabors when they knew that Gabors had different orientations. To avoid such a voluntary strategy shift, we tested many levels of external noise interleaved within the same block so

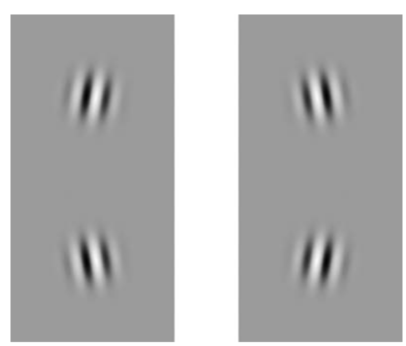

Figure 6. Gabor pairs used in the second experiment. Each sample was composed of two Gabors oriented in opposite orientations relative to vertical, i.e., the upper Gabor could be tilted clockwise and the lower one counterclockwise from vertical (right response) or vice versa (left response). On each trial either one or two of these pairs would be presented (Figure 7).

observers did not know a priori whether the samples had the same orientation or not. As a result, they could not voluntarily change their processing strategy as a function of the external noise level.

Finally, we varied the size of the Gabors (which should affect the precision of the sample estimates, i.e., early noise) to test for effects of late noise. In the first experiment, we found a manipulation (set size) that affected performance only in high noise. In the second experiment, we looked for the opposite effect by also manipulating the Gabor size (i.e., width of the spatial envelope). Increasing the Gabor size is expected to increase the precision of the sample estimates but would have little impact on the averaging process. For the change in Gabor size we used ($SD = 0.23$ and 0.33 dva), we expect a change in performance in low external noise but not in high external noise. But if the averaging advantage in low noise is hidden by late noise (e.g., decisional noise), then the advantage of more precise sample estimates (i.e., less early noise) due to increased Gabor size should also be hidden: performance should be independent of the Gabor size. In the second experiment, we therefore measured the impact of set size and Gabor size.

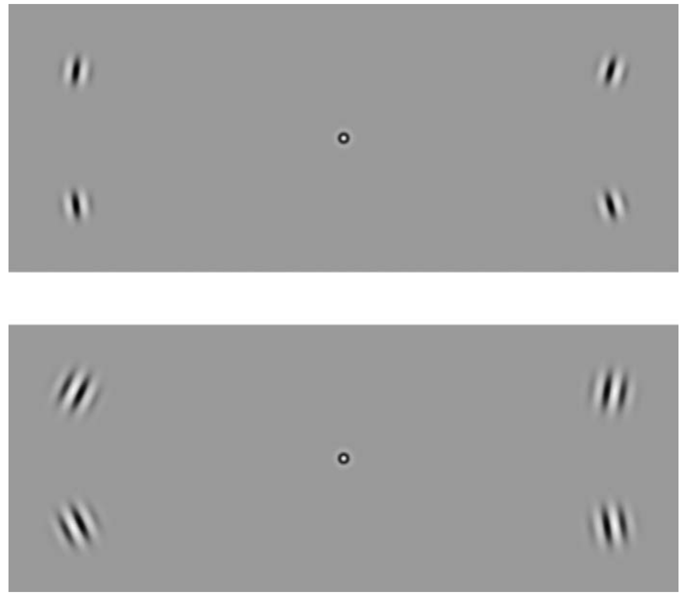

Figure 7. Stimulus examples of the second experiment. Each stimulus was composed of either two samples presented 8° to the left and to the right of fixation (as presented here) or one sample randomly presented 8° to the left or to the right of fixation (not shown). The standard deviation of the Gaussian window of each Gabor was either 0.23° (top) or 0.33° (bottom).

## Method

### Observers

Two naïve observers and an author participated in this experiment. They all had normal or corrected-to-normal vision.

### Stimuli

The stimulus was composed of one or two samples, which were displayed 8° of visual angle to the left and right of fixation (Figure 7). Each sample was composed of two Gabors positioned 2° above and below fixation level (i.e., horizontal line crossing fixation). Both Gabors were symmetrical relative to the horizontal axis: they had the same phase and opposite orientations. Observers were instructed to determine if the sample(s) was(were) "pointing" to the left (Figure 6 left) or to the right (Figure 6 right). Specifically, when two samples were presented (Figure 7), observers were instructed to estimate the orientation difference between top and bottom Gabors of each pair (i.e., angle of the "pointer") before averaging these estimates. Note that observers could have ignored the instructions and used a different strategy. They could have averaged the orientations of the two top Gabors and then compared it to the averaged orientation of the two bottom Gabors. But even if observers used such a strategy, it would not compromise the main target of the present experiment, which was to null the impact of an orientation bias common to all samples. Besides the number of samples (one or two), we also manipulated the Gabor size: the standard deviation of the Gaussian spatial window was either 0.23 or 0.33 dva (Figure 7). All other stimulus settings were identical to the previous experiment.

### Procedure

For each of the four conditions (2 set sizes × 2 Gabor sizes), orientation discrimination threshold was measured under 8 noise levels ($SD$ of the Gaussian distribution of 0°, 0.25°, 0.5°, 1°, 2°, 4°, 8° and 16°), which were pseudo-randomly interleaved within the same block. Each condition was blocked and tested three times in a pseudo-random order, resulting in a total of nine blocks. The remaining settings were identical to the first experiment.

### Data fitting

Orientation discrimination threshold as a function of external noise level has a stereotypical hockey-stick function in log-log coordinates gradually shifting between a flat asymptote (slope = 0) in low noise and a rising asymptote with a slope of 1 in high noise (Figure 2). The flat asymptote represents the orientation discrimination threshold in absence of noise whereas the rising asymptote represents the orientation discrimination threshold relative to external noise level in high noise.

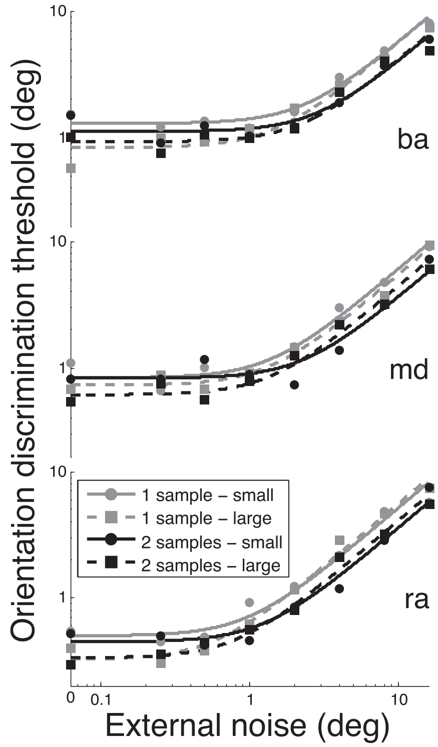


Figure 8. Mean orientation discrimination threshold as a function of the *SD* of the pdf (i.e., external noise) for the three observers. For each condition and each subject, the data were fitted using Equation 7 (which is mathematically equivalent to the averaging model).

For each condition and each subject, orientation discrimination threshold as a function of external noise level ($c(n)$) were fitted using the following function (Allard & Cavanagh, 2011):

$$c(n) = \sqrt{a_{low}^2 + (a_{high}n)^2} \tag{7}$$

where $a_{low}$ represents the orientation discrimination threshold in absence of external noise (i.e., low-noise asymptote) and $a_{high}$ represents the orientation discrimination threshold relative to the external noise level required to discriminate the target in high noise (i.e., high-noise asymptote). Consequently, there was a free parameter having a significant impact only in low noise ($a_{low}$) and another having a significant impact only in high noise ($a_{high}$). Note that this function is mathematically equivalent to the averaging model, which has a parameter having a significant impact only in low noise (i.e., internal noise) and another affecting thresholds in both low and high noise (averaging efficiency).

## Results and discussion

Data were well fitted with Equation 7 (Figure 8) for the four testing conditions. Based on these fits, Figure 9 shows the set-size effect (one vs. two samples) and the Gabor-size effect (spatial window of 0.23 vs. 0.33 dva). Reducing the Gabor size from 0.33 dva to 0.23 impaired performance in low noise by a factor of about 1.35 and had little impact in high noise (Figure 9, left). Conversely, reducing the number of samples from 2 to 1 impaired performance in high noise by a factor of about 1.4 (or reduced averaging efficiency by a factor of 2.0) and had little impact in low noise (Figure 9, right).

The double dissociation suggests that different processing strategies underlie relative mean orientation discrimination in low and high noise. Critically, the absence of averaging efficiency in low noise despite the substantial averaging efficiency in high noise cannot be explained by an orientation bias such as a noisy internal vertical reference or by a processing strategy shift depending on the block condition since all noise levels were interleaved. Furthermore, the absence of averaging efficiency in low external noise cannot be explained by late noise occurring after the averaging process (e.g., decisional noise). If it was, then the effect of Gabor size would have been limited as well. Indeed, reducing early noise by improving the precision of the sample estimates or reducing the impact of early noise by averaging would have a negligible impact if late noise was greater than early noise. But improving the precision of sample estimates by increasing the Gabor size was found to improve performance, which shows that performance was limited by early noise (i.e., the precision of sample estimates), not by late noise.

Given that early noise limited performance in low noise and that observers were efficiently averaging across samples in high noise, the noise-invariant processing assumption implies that the averaging process should improve performance in low noise. This result was not observed. Nevertheless, we will briefly consider one further possibility that might account for the absence of an averaging advantage that we found with several identical samples in low noise. Specifically, the averaging process itself might add noise, counteracting the improvement from averaging when there is little external noise. Since similar thresholds were obtained with one and two samples, this "averaging noise" would have to happen to be slightly lower than the early noise (about 0.7×). If the averaging noise was much lower than the early noise, then it would have no significant impact, and threshold with two samples would be better than with 1 sample. If the averaging noise was equal or higher than the early noise, then thresholds with two samples would be worse than with one sample. Given that averaging noise would be independent of the Gabor size and that early noise was lower with larger Gabors (we found a substantial Gabor-size effect with 1 sample), averaging noise must (a) reduce the Gabor-size effect (i.e., there should be a smaller Gabor-size effect with two compared to 1

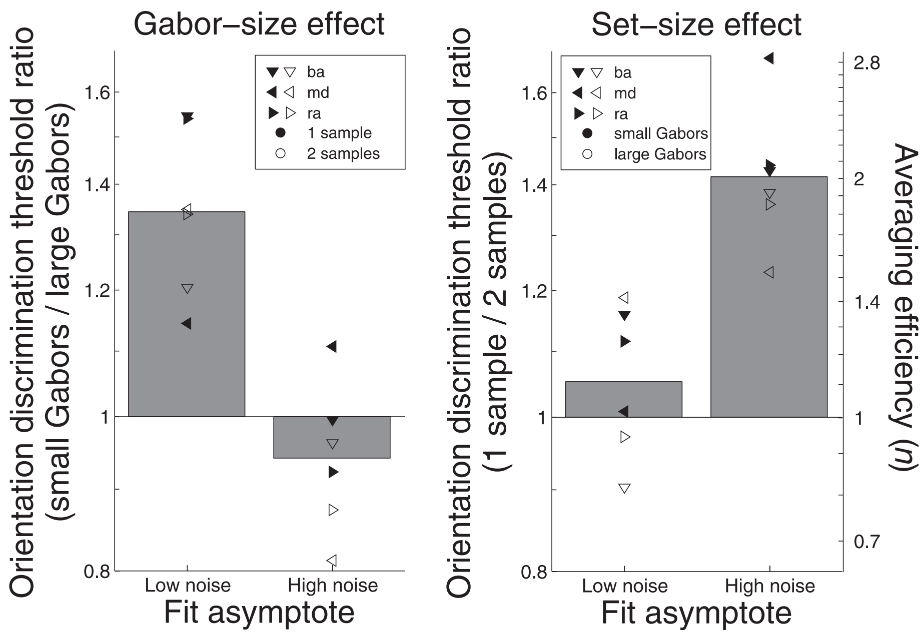


Figure 9. Mean Gabor-size (left) and set-size (right) effects presented as orientation discrimination threshold ratios for both the low ($a_{low}$) and high ($a_{high}$) noise asymptote fits in Figure 8. Individual Gabor-size effects (left) were averaged across both one and two samples (filled and open symbols, respectively). Individual set-size effects (right) were averaged across both small and large Gabors (filled and open symbols, respectively). Note that the set-size effects (right) can be presented as threshold ratios ($O_1(N_{ext})/O_2(N_{ext})$) or averaging efficiencies (Equation 6).

sample) and (b) reduce the set-size effect especially when the early noise is low (i.e., there should be a smaller set-size effect with large compared to small Gabors). We did not find either of these effects in our data (Figure 9), but we do not have the statistical power to rule out this hypothesis. However, we consider the hypothesis that averaging both improves (reduces the impact of early noise) and impairs (adds extra noise of slightly lower than early noise) performance by a similar amount to be less parsimonious than the hypothesis that observers were simply not averaging apparently identical samples.

Furthermore, this hypothesis would require the existence of *additive* averaging noise. Additive noise would affect performance only in low external noise, as opposed to multiplicative noise that would affect performance equally in low and high noise. If the averaging process has intrinsic additive noise, then it would be theoretically possible to find a condition in which the additive averaging noise is higher than the early noise, resulting in a better performance with one rather than two samples. As a result, an observer should be able to correctly discriminate each of two samples as slightly tilted clockwise but would be at chance when reporting their average. We are not aware of any evidence supporting this counterintuitive effect. We consider this possibility as unlikely and believe that the averaging noise is multiplicative, not additive. We therefore conclude that the absence of averaging efficiency in low noise was observed because observers were not averaging apparently identical samples, not because the averaging process has intrinsic additive noise that happens to be of slightly lower than the early noise. Given that observers were efficiently averaging in high noise, the fact that observers were not averaging in low noise implies that different processing strategies underlie relative mean orientation discrimination in low and high noise.

## General discussion

In the current study, observers were able to efficiently average many samples in high noise but not in low noise. These results are inconsistent with the averaging model typically used with the voluntary averaging paradigm (Figure 1b), which assumes that the averaging efficiency is independent of the noise level, i.e., whether the dominating noise source limiting sample precision estimates comes from the observer (i.e., internal) or from the stimulus (i.e., external). We therefore conclude that the averaging model (Figure 1) cannot be used to model our results for the two tasks used in the present study.

Previous studies have based their interpretation on the averaging model and assumed that the averaging process had the same efficiency in low and high noise. Thus, our results showing different averaging efficiencies in low and high noise questions previous interpretations based on this assumption. For instance, as mentioned in the Introduction, Dakin et al. (2009) used a voluntary averaging paradigm to compare the effect of attentional load and crowding on a mean orientation discrimination task (similar to the task in our first experiment). On the one hand, they found that crowding only had an effect in low noise so they concluded that crowding affects the precision of sample estimates (i.e., internal noise). On the other hand, attentional load was found to affect performance in both low and high noise leading them to conclude that attentional load affected averaging efficiency assumed to be equally effective in both low and high noise. They therefore concluded that attentional load and crowding had dissociable effects: crowding affected sample precision estimates without affecting averaging efficiency, and attentional load affected averaging efficiency without affecting sample precision estimates. They argued that this "double dissociation" between attentional load and crowding suggests that crowding does not reflect an attentional limit as suggested by others (He, Cavanagh, & Intriligator, 1996; Intriligator & Cavanagh, 2001). However, Dakin and colleagues (2009) inferred that the attentional load effect in low noise was due to a lower averaging efficiency based on the averaging model in which the averaging efficiency is assumed to be the same in low and high noise. This interpretation is inconsistent with our results showing a significant averaging efficiency only in high noise. If the attentional load effect in low noise was not due to a lower averaging efficiency, then it could be due to poorer fine orientation discrimination (i.e., lower precision of the sample estimates). Thus, if the averaging process has different efficiencies in low and high noise, their results no longer support a double dissociation between crowding and attentional load. Their results could rather be interpreted as: (a) crowding and attentional load both affect fine orientation discrimination (i.e., precision of sample estimates), which only affects performance in low noise, and (b) attentional load alone affects averaging efficiency, which only affects performance in high noise. Thus, our results showing no significant averaging efficiency in low noise calls into question previous interpretations based on the averaging model.

In the second experiment, we observed a double dissociation between voluntary averaging in low and high noise: varying Gabor size affected performance only in low noise and varying the number of samples affected performance only in high noise. Based on this double dissociation, we conclude again that different processing strategies underlie voluntary averaging in low and high noise. In high noise, observers must have been averaging across the two samples since their performances were better than the ideal performance given one sample. But in low noise, we found no evidence of an efficient averaging process: similar performances were observed with one and two samples. This suggests that the averaging process operating in high noise was not operating in low noise. In low noise, we suggest that observers were not averaging across the two samples because they appeared to be identical (presumably due to a thresholding mechanism, Morgan, et al., 2008; Ross & Burr, 2008). Thus, although it is theoretically possible for an observer to voluntarily average identical samples, averaging identical samples is trivial and does not require the same resources as averaging different samples. Since we found no evidence that observers averaged apparently identical samples (i.e., in low noise), we conclude that they did not. This absence of averaging in low noise implies that the processing strategy changed depending on the external noise level, violating the noise-invariant processing assumption underlying voluntary-averaging paradigms.

Analogously, we could speculate that a fine orientation discrimination process is only engaged in low noise. Fine orientation discrimination, which presumably requires more resources than coarse orientation discrimination (Pelli, Palomares, & Majaj, 2004), would improve performance in low noise but not in high noise when performance does not depend on the precision of sample estimates. Estimating orientation with a high level of precision is certainly crucial in low noise when all Gabors are close to vertical. But in high noise when the Gabors are generally far from vertical, estimating the orientation of each with a high level of precision is useless. For instance, discriminating whether a sample is tilted 1° to the left or to the right of vertical can be useful when all samples are near vertical, but discriminating whether a sample is tilted 24° or 26° to the left has negligible impact in high noise. Thus, we consider it likely that fine orientation discrimination is only engaged in low noise as there is no need for much precision in high noise. This is only speculation, though, since the performance in high noise would not be affected whether fine orientation discrimination was engaged or not.

Note that our data showing substantial averaging efficiency only in high noise but negligible averaging efficiency in low noise are analogous to those reported by Solomon (2010) for a different orientation discrimination task. But our interpretations differ drastically. We conclude that the voluntary-averaging process only operates in high noise; and Solomon concluded instead that the same averaging process operates in both low and high noise, but the averaging efficiency does not affect performance in low noise due to late internal noise, occurring after the averaging process. The main difference between our task in Experiment 1 and his task in his experiments 2 and 3 is that his task consisted in determining whether the mean orientation of a first set of Gabors was tilted clockwise or counterclockwise from the mean orientation of a second set of Gabors. Thus, the overall orientation estimation of the first set of Gabors was used as an internal reference to estimate the mean orientation of the second set of Gabors. As we mentioned in the discussion of the first experiment, such a comparison with an internal reference could introduce late noise that would null the averaging efficiency in low but not in high noise. Although this interpretation is a plausible explanation of Solomon's data, it is not consistent with the results of our second experiment in which no internal reference was required. More generally, the absence of averaging efficiency in low external noise cannot be explained by late noise occurring after the averaging process. If mean orientation discrimination in low external noise were limited by late noise, then performance would be independent of both the averaging efficiency and the precision of the sample estimates (i.e., early noise). But the results of the second experiment showed that performance depended on the precision of the sample estimates manipulated by varying the Gabor size. This effect shows that mean orientation discrimination in low noise were limited by early noise (i.e., precision of sample estimates) and thereby rules out the possibility that the absence of averaging efficiency in low noise was due to late noise. We therefore conclude that, at least under some conditions, the processing strategy changes depending on the external noise level.

Our conclusion that the noise-invariant processing assumption does not hold for a voluntary-averaging task is consistent with our previous claim that the processing strategy can change as a function of external noise level (Allard & Cavanagh, 2011). Our previous study on luminance detection and the present study on mean orientation discrimination both show that an experimenter cannot assume a priori that the processing strategy is the same in low and high noise. This noise-invariant processing assumption is so widely accepted that it is often not explicitly stated; experimenters vary the noise level (e.g., contrast of luminance noise or orientation variance) without wondering if it affects the processing strategy. Showing that this assumption can be violated, here for the second time, implies that experimenters using external noise paradigms must consider the possibility that the processing strategy can change with external noise level.

Nonetheless, the fact that the processing strategy *can* change as a function of external noise level does not necessarily imply that the processing strategy *must* change. In some conditions, processing could be independent of external noise levels, i.e., the same processing strategy could be effective whether the dominating noise originates from the observer (i.e., low external noise) or the stimulus (i.e., high external noise). If the processing strategy does not change with external noise levels, then external noise paradigms could be used to characterize the processing of a given stimulus. Unfortunately, we know no test that can assert that the processing strategy has remained the same whether the dominating noise source comes from the observer (i.e., internal) or the stimulus (i.e., external). Since the noise-invariant processing assumption plays a central role in external noise paradigm, this substantially weakens conclusions based on external noise paradigms.

## Acknowledgments

This research was supported by a FQRNT postdoctoral fellowship to RA and a Chair d'excellence grant to PC.

Commercial relationships: none
Corresponding author: Remy Allard.
Email: remy.allard@umontreal.ca.
Address: Laboratoire Psychologie de la Perception, Paris, France.

## References

Allard, R., & Cavanagh, P. (2011). Crowding in a detection task: External noise triggers change in processing strategy. [Research Support, Non-U.S. Gov't]. *Vision Research, 51*(4), 408–416.

Allard, R., & Faubert, J. (2006). Same calculation efficiency but different internal noise for luminance- and contrast-modulated stimuli detection. *Journal of Vision, 6*(4):3, 322–334. http://www.journalofvision.org/content/6/4/3, doi:10.1167/6.4.3. [PubMed] [Article]

Allard, R., & Faubert, J. (2008). First- and second-order motion mechanisms are distinct at low but common at high temporal frequencies. *Journal of Vision, 8*(2):12, 1–17. http://www.journalofvision.org/content/8/2/12, doi:10.1167/8.2.12. [PubMed] [Article]

Bennett, P. J., Sekuler, A. B., & Ozin, L. (1999). Effects of aging on calculation efficiency and equivalent noise. *Journal of the Optical Society of America, A, Optics, Image Science, & Vision, 16*(3), 654–668.

Dakin, S. C. (1999). Orientation variance as a quantifier of structure in texture. *Spatial Vision, 12*(1), 1–30, doi: 10.1163/156856899x00012.

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of* *the Optical Society of America, A, Optics, Image Science, & Vision, 18*(5), 1016–1026.

Dakin, S. C., Bex, P. J., Cass, J. R., & Watt, R. J. (2009). Dissociable effects of attention and crowding on orientation averaging. *Journal of Vision, 9*(11):28, 1–16, http://www.journalofvision.org/content/9/11/28, doi:10.1167/9.11.28. [PubMed] [Article]

Dakin, S. C., Mareschal, I., & Bex, P. J. (2005). Local. and global limitations on direction integration assessed using equivalent noise analysis 24. *Retrieved UCL, Inst Ophthalmol, Dept Visual Sci.* (pp. 3027–3049). 11-43 Bath St, London EC1V 9EL, UK, 45.

Dosher, B. A., & Lu, Z.-L. (2004). Mechanisms of perceptual learning. In L. Itti & G. Rees (Eds.), *Neurobiology of attention* (pp. 471–476). San Diego, CA: Academic press.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10–90). New York: Wiley.

He, S., Cavanagh, P., & Intriligator, J. (1996, September 26). Attentional resolution and the locus of visual awareness [Letter]. *Nature, 383*(6598), 334–337.

Intriligator, J., & Cavanagh, P. (2001). The Spatial Resolution of Visual Attention. *Cognitive Psychology, 43*(3), 171–216.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*(2), Suppl 2467.

Lu, Z.-L., & Dosher, B. A. (2004). External noise distinguishes mechanisms of attention. In L. Itti & G. Rees (Eds.), *Neurobiology of attention* (pp. 448–453). San Diego, CA: Academic press.

Lu, Z.-L., & Dosher, B. A. (2008, January). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review, 115*(1), 44–82.

Morgan, M., Chubb, C., & Solomon, J. A. (2008). A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision, 8*(11):9, http://www.journalofvision.org/content/8/11/9, doi:10.1167/8.11.9. [PubMed] [Article]

Pardhan, S. (2004). Contrast sensitivity loss with aging: sampling efficiency and equivalent noise at different spatial frequencies. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 21*(2), 169–175.

Pardhan, S., Gilchrist, J., & Beh, G. K. (1993). Contrast detection in noise: a new method for assessing the visual function in cataract. *Optometry & Vision Science, 70*(11), 914–922.

Pelli, D. G. (1981). *The effects of visual noise*. Dissertation., Cambridge University, Cambridge.

Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Visual coding and efficiency*. (pp. 3–24). Cambridge: Cambridge University Press.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America, A, Optics, Image Science, & Vision, 16*(3), 647–653.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12):12, 1136–1169, http://www.journalofvision.org/content/4/12/12, doi:10.1167/4.12.12. [PubMed] [Article]

Ross, J., & Burr, D. (2008). The knowing visual self. *Trends in Cognitive Sciences, 12*(10), 363–364, doi:10.1016/j.tics.2008.06.007.

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision, 10*(14):9, 1–16, http://www.journalofvision.org/content/10/14/19, doi:10.1167/10.14.19. [PubMed] [Article]