Parallel extraction of summary information across multi-element arrays

March 27th, 2013

1 Introduction

Observers are capable of making rapid and accurate judgements of the summary information in an array composed of multiple elements [Ariely 2001]. Whilst some theories have argued that this suggests that the visual system is capable of automatic extraction of statistical information from a visual scene [Cavanagh 2001, Robitaille 2011], others have suggested that serial strategies may play a role [Myczek 2008].

Despite evidence showing that the vision system in humans is able to process at least some information in a parallel way [?], it has been argued that serial mechanisms can actually account for current results during averaging [Myczek 2008]. However, it has been noted that this argument might involve a fallacy because it only takes into account the distribution of the array (but not the perceptual noise) and it focuses on a particular task rather than a more general domain to which the parallel model has been successfully validated [Ariely 2008].

To test whether perceptual averaging occurs in parallel or in series we designed, runned and analysed three different experiments.

2 Methods

2.1 Experiment 1 (Shape categorisation – Variance + Setsize)

In a first experiment, we asked observers to judge the average feature (shape or colour) in a centrally-presented visual array with a variable number of elements (‘squircles’, i.e. shapes that varied continuously from shape to circle).

2.1.1 Participants

We ran the experiment with 10 participants: 3 males and 7 females, between 18 and 25 years old. Participants trained themselves before the actual task.

They were happily paid for their collaboration.

2.1.2 Procedure and materials

A set of squircles (‘squircles’, i.e. shapes that varied continuously from shape to circle) was displayed across a ring centred on the screen. The position of the elements was fixed for each setsize (i.e., for 2, 4 or 8 items).

In this experiment we explicitly manipulated three independent variables:

- ✗ Five levels of the average shape of the array – counterbalanced between negative (square) and positive (circle) values (i.e., nine levels of mean in total).

- ✗ Two levels of variance of the shape of the array – low and high.
- ✗ Three different sizes of the array (i.e., setsize) – 2, 4 and 8.

The experiment was divided into blocks of 25 trials. All trials in a block had same variance and setsize.

The task was designed in the following way (see Figure 1).

On each trial, participants saw a fixation point, followed by a stimulus presented for 200ms. After this time, a circular mask appeared for another 300ms. Participants should then report the category of the average shape of the set – without time constraints.

Categories were reported using the buttons in the mouse – which were counterbalanced across subjects.

A CRT-Screen was used for all of the experiments.

Auditory feedback was provided on a trial-per-trial basis.

Do we need to report the size of the squircles/ring?

2.1.3 Results

Consistently with previous reports [de Gardelle 2011], we found (through ANOVA) a **significant effect of mean** ($F=68.575$, $p<0.000$) **and variance** ($F=7.651$, $p<0.022$) on performance – with better performance corresponding with higher levels of mean and lower level of variance. However, **no significant effect was found for setsize** ($F=0.147$, $p<0.835$) – and the correlation between performance and setsize, though non-significant, is positive.

No interactions were found either.

No effects are found on reaction times (RT).

Should check the anova

2.1.4 Discussion

Similar results have previously been reported on performance. Supporting these findings, in [Robitaille 2011] an effect on setsize is found, with bigger setsizes correspond to better performance. This is directly contradictory with explanations involving serial mechanisms, for which the proportion of the number of items processed over the total number of items (the setsize) should decrease, and so the performance.

An important distinction of the present study is that the **sampling of the array from an underlying distribution is constrained**, such that actual average and variance of the array in the display represent accurately the ones of the underlying distribution – from which the performance is derived. If no such constraint is done, there is a theoretical advantage in estimating the underlying mean from more samples, which can be taken as an explanation for the positive correlation between performance and setsize – as discussed in [Robitaille 2011]. In the current experiment the information brought by each item was inversely proportional to the setsize, and so there was a disadvantage in sampling a subset of items, rather than an advantage in sampling more items for trials with bigger setsizes.

Constraining the sampling in this experiment is a crucial contribution to show that average estimation accuracy doesn't decrease with setsize, supporting parallel models for categorisation of the average of an array of visual stimuli.

2.2 Experiment 2 (Shape categorisation – Duration + Setsize)

Experiment 1 brings evidence that supports parallel mechanisms. If using serial strategies (and having a constrained sampling), or if the number of items that we can process in parallel is lower than the setsize of the array, then performance should decrease with the setsize – which is actually not happening!

However, it might happen that humans have a limited capacity of sampling, but can focus serially on different items if they have the time to do it.

Following this idea, performance should also start decreasing abruptly with the duration of the presentation after a certain threshold – if all items cannot be processed in parallel or during the presentation duration – and we should find an interaction on performance between setsize and the duration the stimuli if they are following a serial strategy.

Thus, we ran a second experiment keeping the same paradigm.

2.2.1 Participants

The experiment was done by 10 happy and healthy participants (9 females and 1 male) between 18 and 26 years old – which were paid for their collaboration.

2.2.2 Procedure and materials

We fixed the variance of the array (to a value mid-way between the two values in the experiment before) and we manipulated mean, duration and setsize independently:

- ✗ Five levels of mean of the shape of the array, counterbalanced between square and circle (i.e., nine levels of mean in total).
- ✗ Four sizes of the array – 1, 3, 6 and 12.
- ✗ Four durations – 100ms, 200ms, 400ms and 800ms.

2.2.3 Results

We found a **main effect of mean** ($F=29.991$, $p<0.000$) **and timing**

($F=6.216$, $p<0.07$), which are positively correlated with performance. A look at the data showed that **subjects are biased for setsize 1** (i.e., when only one item is presented). Thus, there was a main effect for setsize which disappeared after removing this condition.

No significant interactions were found.

2.2.4 Modelling

As an interpretation of these results, we defined a simple model to explain the previous findings. This model extends the one found in [Myczek 2008]. Figure 4a illustrates the principle of the model.

The model has three parameters: the *capacity*, the *duration power* and the *perceptual noise factor*. At each trial, the model selects randomly a number of items/samples equal or lower than its *capacity*. Some perceptual noise (following a normal distribution with null mean) is added to each sample. This perceptual noise has a standard deviation which depends on the *perceptual noise factor* and the *duration power*. Formally

$$x_i = s_i + n_i \text{ for } i<c,$$

$$\text{with } n_i = N(0,1) * PNF * D^{DP}$$

where c is the *capacity*, N(0,1) is sample from a normal distribution with mean 0 and variance 1, PNF is the *perceptual noise factor*, D is the duration of the stimulus (in seconds) and DP is the *duration power*. The output of the model is a categorisation over the average across all x values:

$$y = \text{sign}(\text{mean}(x_i)), \text{ for all } i$$

Figure 4b show the result of performance for this model.

Observe that the *capacity* of the model allows to adjust how performance decreases with setsize – particularly for small sets. In opposition, the *perceptual noise factor* allows to adjust how performance increases with setsize – particularly for big sets. The combination of both parameters gives rise a U-shape performance across setsizes.

The absence of interactions between setsize and duration supports as well the fact that statistical summary of vision perception in humans happens in a parallel fashion.

2.2.5 Discussion

Since performance in humans starts to decrease for setsizes 6 and 12, we deduce that the capacity of the model should be located between 6 and 12 items as well – i.e., humans should be able to process, theoretically, at least a minimum of 6 items in parallel.

On the other hand, the *duration power* adjusts the slope of performance across durations.

2.3 Experiment 3 (Shape estimation – Duration + Setsize)

Experiments 1 and 2 bring evidence for a parallel processing of visual stimuli during statistical categorisation of arrays.

2.3.1 Participants

The experiment was done by 9 participants (5 females and 4 males) between 22 and 32 years old – which were paid for their collaboration.

2.3.2 Procedure and Materials

In experiment 3, we replicated experiment 2 asking for an estimation of the average – instead of a categorisation. This would allow us to run more sophisticated analysis over the results found. We also used shorter durations for the stimulus:

- ✗ Four durations – 50ms, 100ms, 200ms and 400ms.

In this case, participants were reporting the estimation of the mean by making a choice between 9 different fixed alternatives. At the end of each trial, they could select one between nine different 'squircles' which were shown simultaneously and which corresponded to the 9 different values of mean used in the underlying distributions (five levels of mean, for positive and negative values, including the 0). The auditive feedback was replaced by a visual one, a white square that highlighted the correct response at the end of the trial.

2.3.3 Results

Since this is now an estimation task (not a categorisation one), performance is given by the error between participants' estimation of the average and the actual value.

Multiple bias were found during shape estimation – towards square and towards neutral values. This bias could accurately be characterised as a linear function of the value of the underlying mean.

We ran an ANOVA over the absolute error of the estimation. In both cases, **main effects were found for mean** ($F=6.601$, $p<0.008$), setsize ($F=4.29$, $p<0.043$) and duration ($F=8.034$, $p<0.002$).

Same effects were found over performance (% correct over the 9 possible choices).

We also found a significant **interaction between mean and setsize** ($F=4.29$, $p<0.043$).

Analysis over linear-regression fittings

2.3.4 Modelling

2.3.5 Discussion

2.4 Experiment 4 (Two-dimensional categorisation – Variance + Setsize)

In a fourth experiment, we wanted to manipulate the perceptual noise – which is associated with each individual item rather than with the whole set. The hypothesis was that, by increasing the perceptual noise, there would be an advantage of sets with more items.

2.4.1 Participants

The experiment was done by 16 participants (12 females and 4 males) between 18 and 33 years old – which were paid for their collaboration.

2.4.2 Procedure and Materials

We used the same design as experiment 1. However, this time participants were asked to categorise following two categories which depended on both shape and colour – e.g., category 1 corresponding to red and circle, or blue and square; category 2 corresponding to red and square, or blue and circle. This mapping of the values is explained in figure 6.

Mean, setsize and variance were manipulated (with the same conditions as in experiment 1). The sampling for the value of all items in the array was constrained as in all of the experiments before. Given the value associated to each item, both colour and shape were derived from a uniform distribution of all possible solutions.

2.4.3 Results

The task was pretty difficult. Performance was a lot lower (69.2% correct in experiment 4, compared to 72.8% in experiment 1). Interestingly, we found effects in mean ($F=48.743$, $p<0.001$) and setsize ($F=10.696$, $p<0.001$) on performance but not of variance ($F=0.333$, $p<0.544$).

No interactions were found.

No effects were found for reaction times either.

Should check the anova

See figure 7.

2.4.4 Discussion

Results confirm our hypothesis

Figures

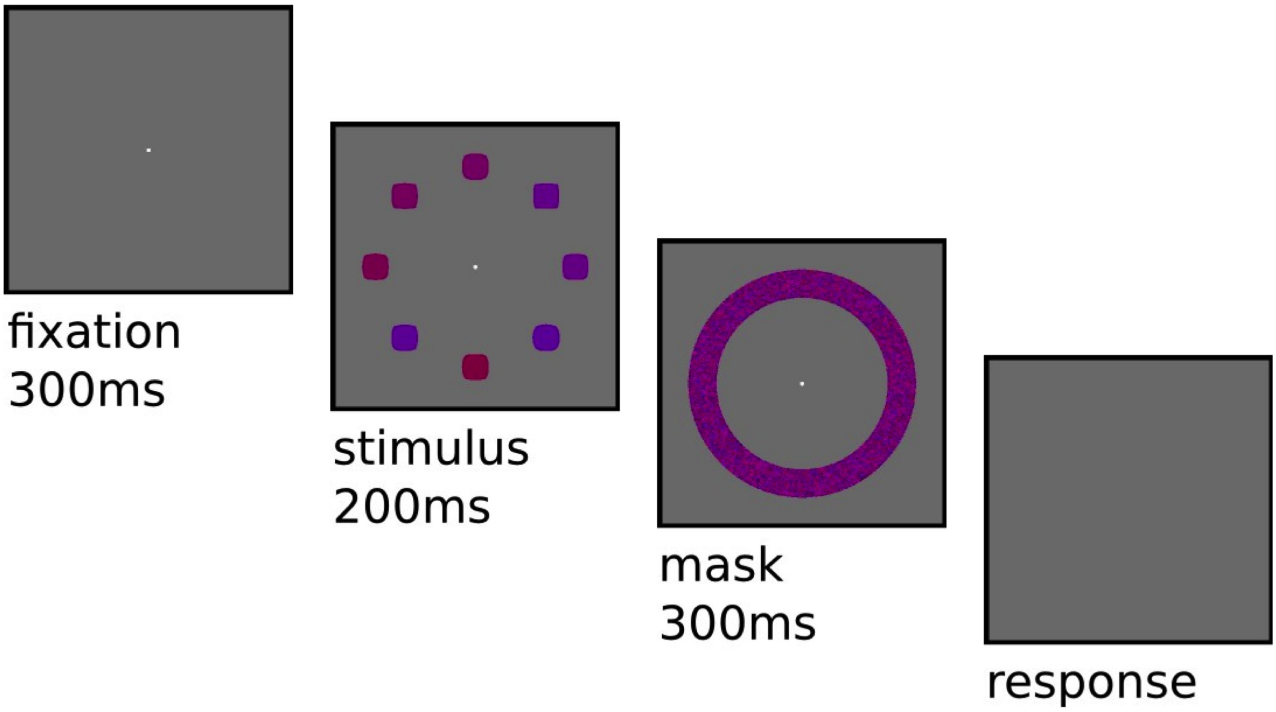

Figure 1. Task design

On each trial, participants first see a fixation point where they have to keep looking during the whole trial. After 300 ms, the stimulus appears – a set of 8 'squircles' with different shape and color. 'squircles' are a continuum of shapes between perfect square and circle. Colour varies from red to blue. After 200 ms, a noisy mask is presented. Finally, participants have to report the category for the average shape of all stimulus – i.e., square-ish or circle-ish. There's no limit in time to respond. An auditive feedback is given at the end of the trial (for correct/incorrect).

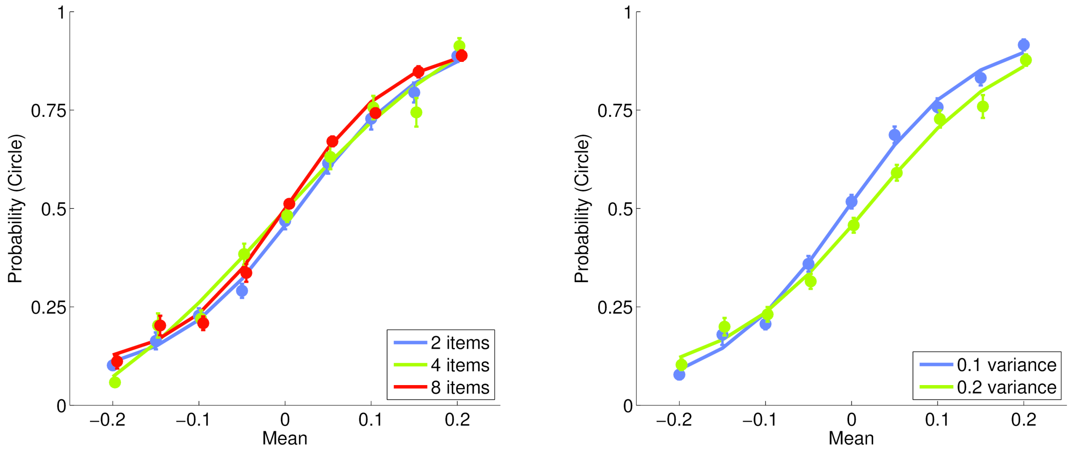

Figure 2a. Psychophysical curves (Experiment 1)

Figure 2a shows the probability of reporting the category 'Circle' for each independent variable (setsize, left; and variance, right).

Circles show the actual data collected. Error bars represent standard deviations across subjects. Lines are fittings using a linear transformation of the sigmoid function (4 free parameters).

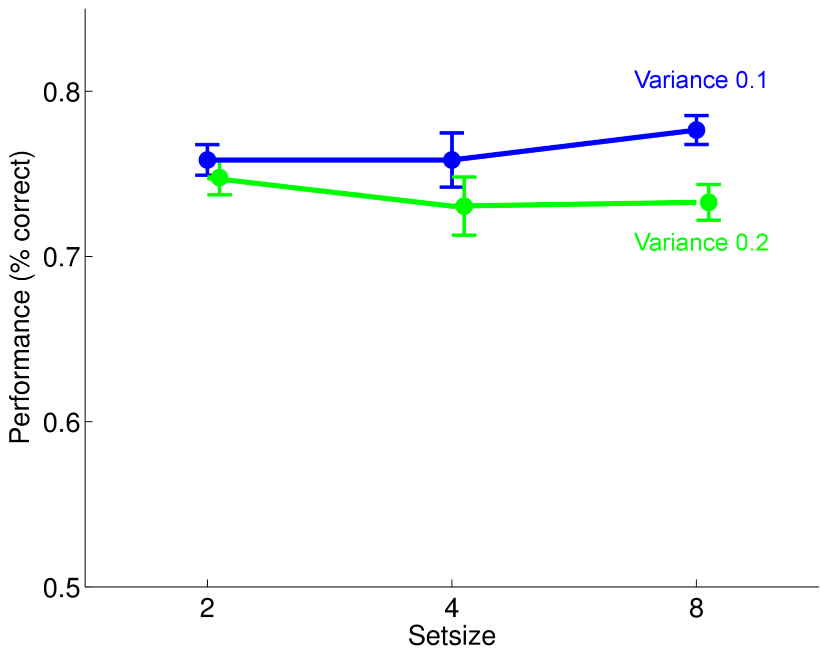

Figure 2b. Performance (Experiment 1)

Figure 2b shows the average performance (% correct) across subjects, for both independent variables. There's no significant effect of setsize on performance. Performance significantly decreases with variance. No significant interactions were found.

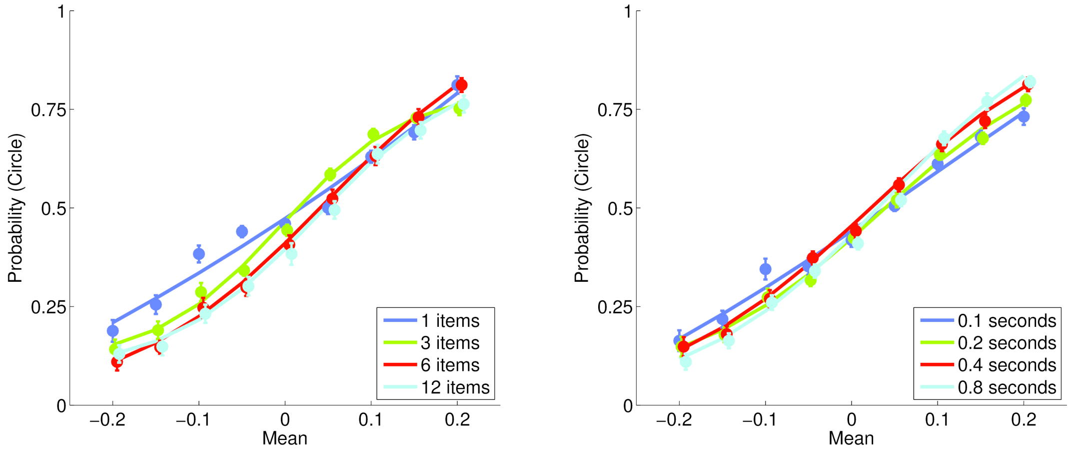

Figure 3a. Psychophysical curves (Experiment 2)

Figure 3a shows the probability of reporting the category 'Circle' for each independent variable (setsize, left; and duration, right).

Circles show the actual data collected. Error bars represent standard deviations across subjects. Lines are fittings using a linear transformation of the sigmoid function (4 free parameters).

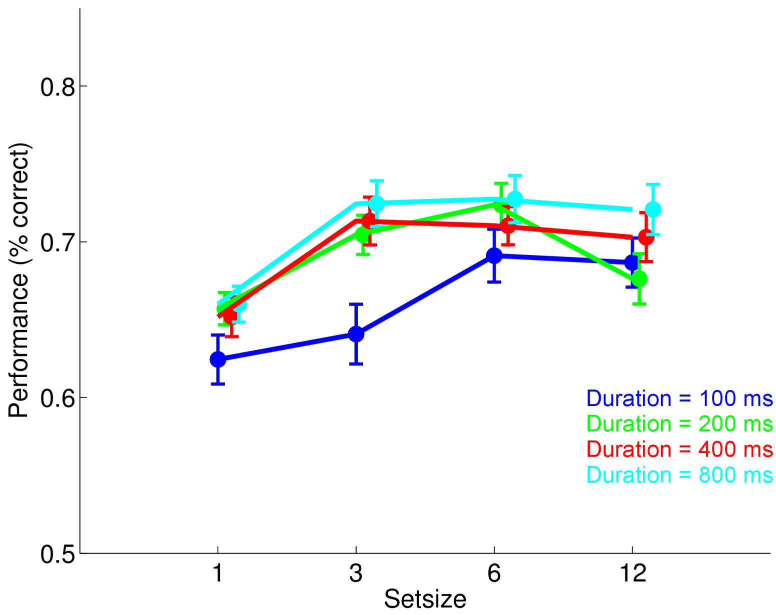

Figure 3b. Performance (Experiment 2)

Figure 3b shows the average performance (% correct) across subjects, for both independent variables (setsize and duration).

There's a significant effect on setsize exclusively driven by the first condition (1 item), with performance increasing with setsize.

Another main effect was found for duration of the stimulus presentation. Performance increased with duration.

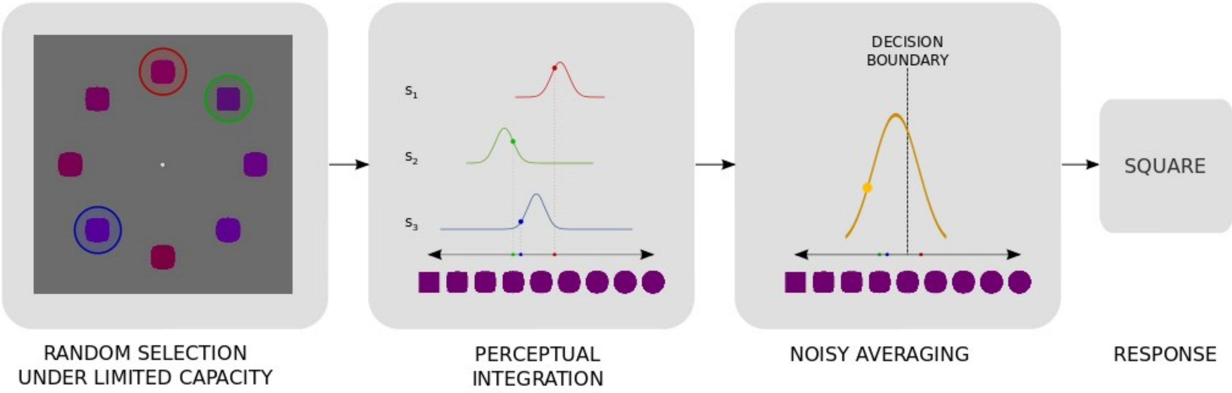

Figure 4a. Diagram of the model
This diagram explains the dynamic of the model. For simplicity's sake, we're using a model with a capacity of 3.

At each trial, the model samples randomly a subset of elements (smaller or equal to its capacity).
A gaussian noise (the *perceptual* noise – which is dependent on the duration of the stimuli) is added to each sample independently.
The average value of the resulting samples is computed. Another gaussian noise (the *policy noise* – independent on the duration) is added to the resulting value.
A category is finally selected by comparing the latter value with an optimal decision boundary.

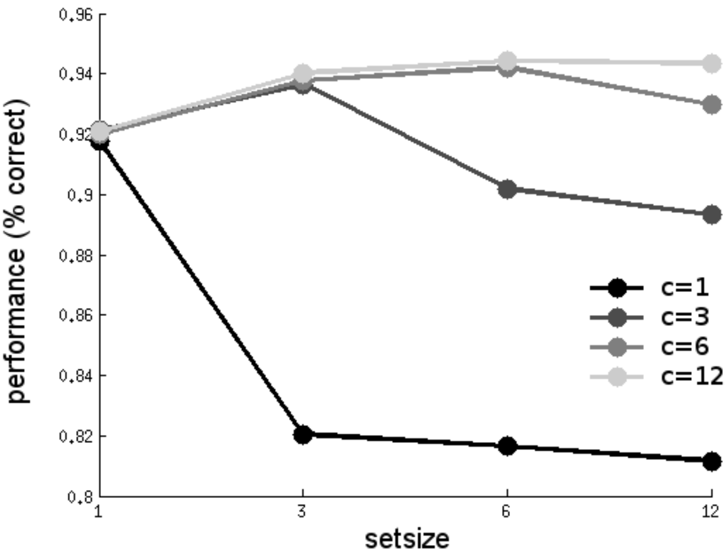

Figure 4b. Performance of the model (experiment 2)
Figure 4b plots average performance (% correct) across setsize of multiple instantiations of the model with different capacities. Inter-individual differences haven't been taken into account. Observe that performance decreases when the number of items to be processed is over the capacity.

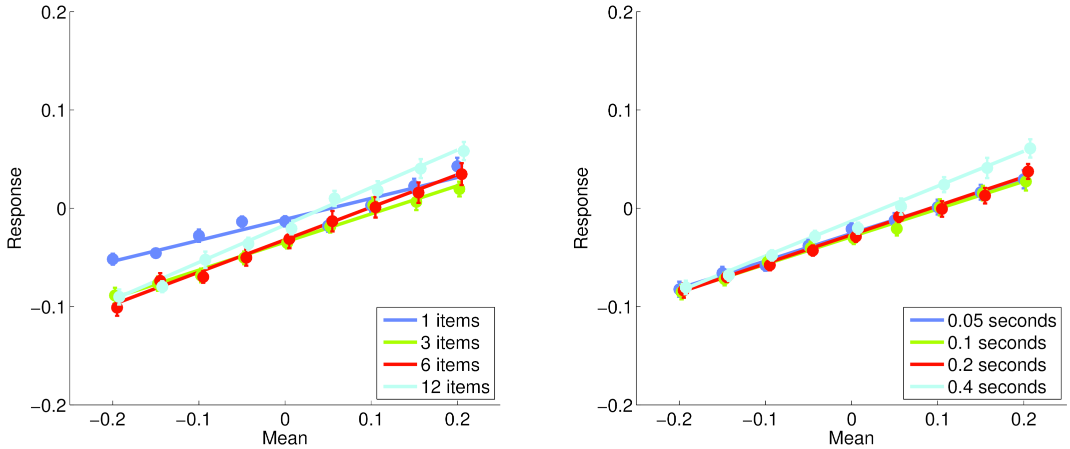

Figure 5a. Estimation response (Experiment 3)
Figure 5a shows the average estimation reported for each independent variable (setsize, left; and duration, right).
Dots show the actual data collected. Error bars represent standard deviations across subjects. Lines are fittings using a linear regression.

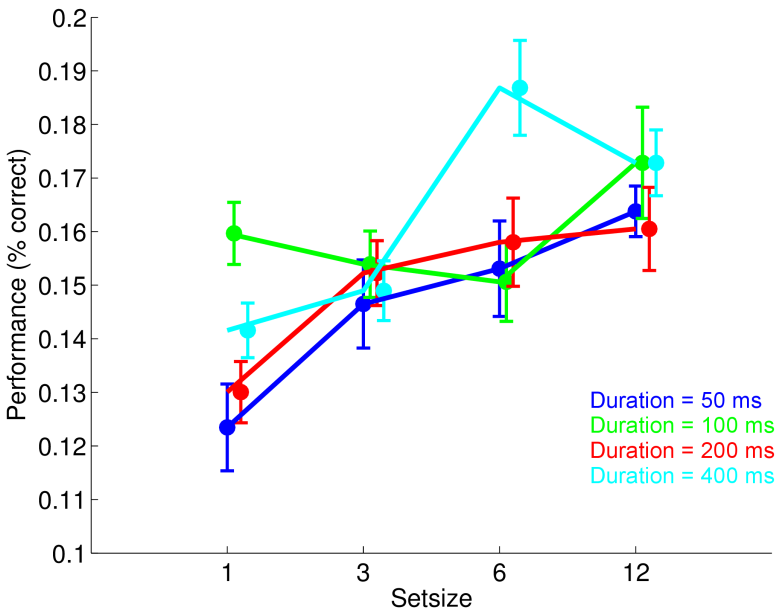

Figure 5b. Performance (Experiment 3)
Figure 5b shows the average performance (% correct) across subjects, for both independent variables (setsize and duration).

Significant effects where found for both setsize and duration, with absolute error decreasing with increasing setsize and/or duration. No significant interaction was found.

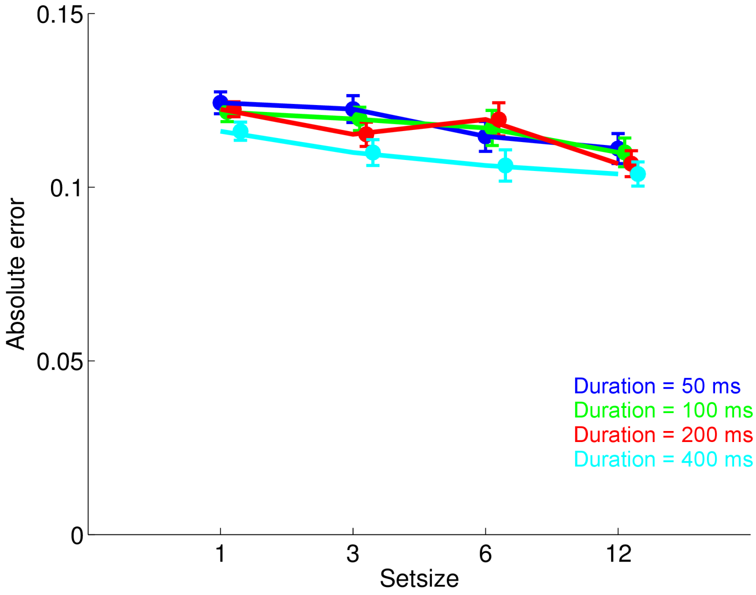

Figure 5c. Absolute error (Experiment 3)
Figure 5c shows the average of the absolute error (value reported minus the actual value) across both setsize and duration.

Significant effects where found for both setsize and duration, with absolute error decreasing with increasing setsize and/or duration. No significant interaction was found.

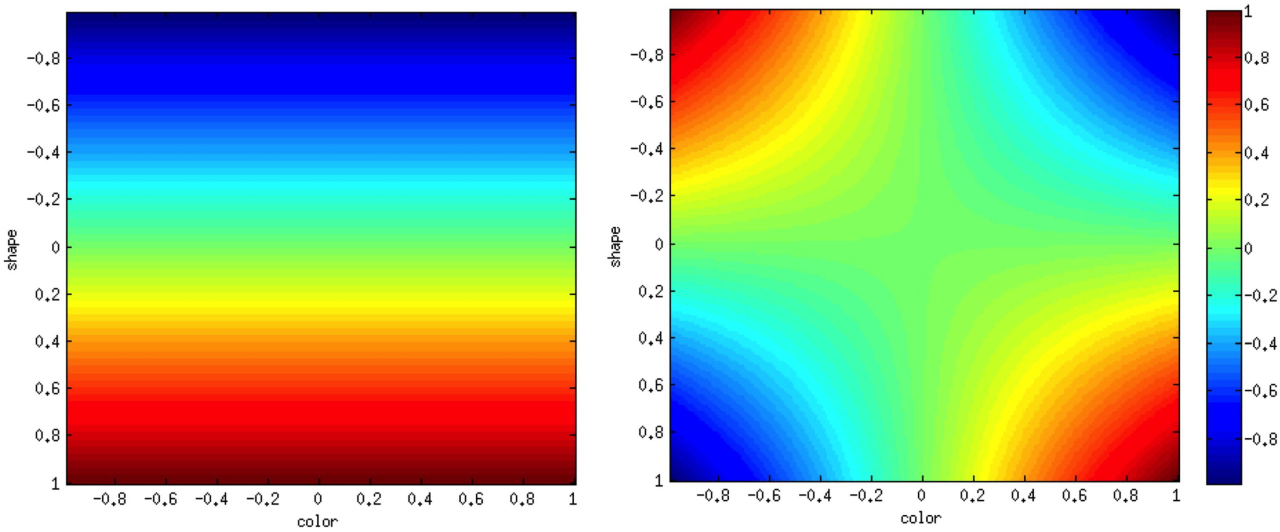

Figure 6. Value-spaces for shape (experiments 1, 2 and 3) and two-dimensional (experiment 4)

The heat-maps above represent the value-spaces for experiments 1,2,3 (left) and 4 (right). The value of each item is associated with a value in the range between -1 and +1 and is evaluated as a function of its colour – red to blue – and its shape – square to circle (see the colour-bar on the right side). Participants are asked to report the average value of the whole set.

In experiments 1, 2 and 3, the value of each item is independent on the colour – so that they're asked to report the average value shape.

The value of each item satisfies the following equation:

$$V_i = S_i$$

In experiment 4, the value of each item is dependent on the colour and the shape of each item and both dimensions interact with each other. Participants are asked, for instance, to report category 1 when items tend to be in average reddish and square-ish, or blueish and circle-ish; report category 2 if not.

The value of each item satisfies the following equation:

$$V_i = C_i * S_i$$

where V_i is the value associated to the i-th item,

C_i is the colour of the i-th item,

and S_i is the shape of the i-th item.

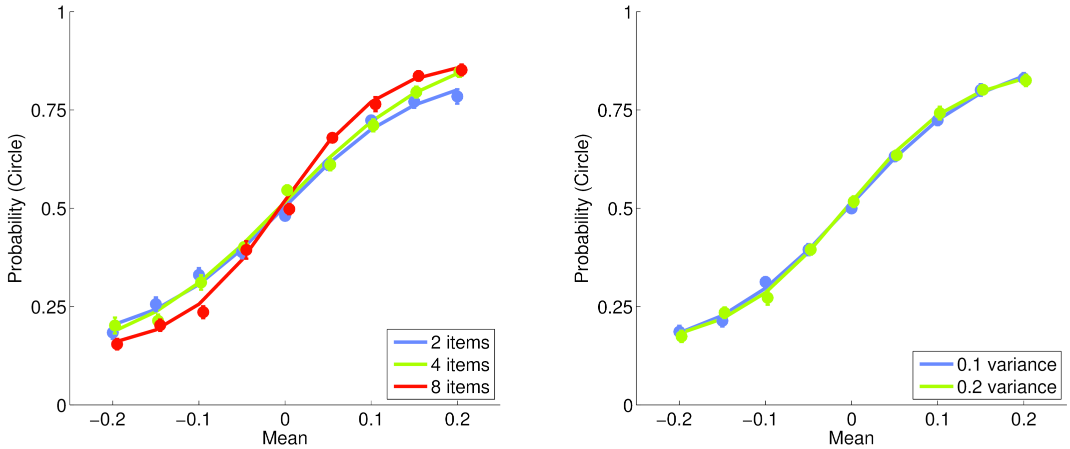

Figure 7a. Category response (Experiment 4)
Figure 7a shows the average category reported for each independent variable (setsize, left; and variance, right).
Dots show the actual data collected. Error bars represent standard deviations across subjects. Lines are fittings using a linear transformation of the sigmoid function (4 free parameters).

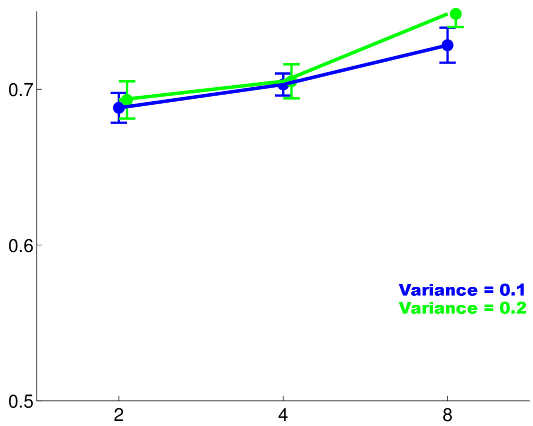

Figure 7b. Performance (Experiment 4)
Figure 7b shows the average performance (% correct) across subjects, for both independent variables (setsize and variance).

A significant effect was found for setsize, with performance increasing with increasing setsize driven by the ss=8 condition. No significant effect was found for variance. No significant interaction was found.

Bibliography

1. Ariely, D. (2001) Seeing sets: Representation by statistical properties.
2. Cavanagh, P. (2001). Seeing the forest but not the trees.

3. Chong, S.C. & Treisman, A. (2003). Representation of statistical properties.
4. Chong, S.C. & Treisman, A. (2004). Statistical processing: computing the average size in perceptual groups.

5. Myczek, K. & Simons, D.J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size.
6. Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations?
7. Chong, S. C., Joo, S. J., Emmanouil, T.-A., & Treisman, A. (2008). Statistical processing: not so implausible after all.

8. de Fockert, J.W. & Marchant, A.P. (2008). Attention Modulates set representation by statistical properties.
9. de Fockert, J.W. & Marchant, A.P. (2009). Priming by the mean representation of a set.

10. de Gardelle, V. & Summerfield, C. (2011). Robust averaging during perceptual judgment.
11. Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets.