

How the deployment of attention determines what we see

**Anne Treisman
Princeton University**

Address correspondence to

Anne Treisman

Psychology Department

Princeton University

Princeton

NJ 08544-1010

e-mail: treisman@princeton.edu

Abstract

Attention is a tool to adapt what we see to our current needs. It can be focused narrowly on a single object or spread over several or distributed over the scene as a whole. In addition to increasing or decreasing the number of attended objects, these different deployments may have different effects on what we see. This chapter describes some research both on focused attention and its use in binding features, and on distributed attention and the kinds of information we gain and lose with the attention window opened wide. One kind of processing that we suggest occurs automatically with distributed attention results in a statistical description of sets of similar objects. Another gives the gist of the scene, which may be inferred from sets of features registered in parallel. Flexible use of these different modes of attention allows us to reconcile sharp capacity limits with a richer understanding of the visual scene.

Perception comprises a range of different ways of informing ourselves about the environment for a variety of different purposes including understanding, recognition and prediction, aesthetic appreciation, and the control of action. One important tool adapting the visual system to different perceptual tasks is the set of control systems that we call attention. Attention can be allocated to different aspects of the environment and in different ways, ranging from the focused analysis of local conjunctions of features to the global registration of scene properties. I will describe some research exploring the effects of these different ways of allocating attention to displays with multiple stimuli.

General framework

I will start by illustrating the framework I use to help me think about the results. Figure 1 shows a first rapid pass through the visual system before attention is brought to bear, or when attention is focused elsewhere. Visual stimuli are initially registered as implicit conjunctions of

features by local units in V1 that respond to many of their properties. At this level they function as filled locations in a detailed map of discontinuities, although of course the information about their properties is implicitly present to be extracted by further processing. Contingent adaptation effects (McCollough 1965) are probably determined at this early level, since they are retinotopically determined, and independent of attention (Houck and Hoffman, 1986).

Figure 1 about here

Specialized units then respond selectively to different “features”, and form feature maps in various extra-striate areas. A ‘red’ unit, for example, would pick up from any V1 detectors that respond to red stimuli in their receptive fields, regardless of the other properties to which they are tuned. The feature maps retain implicit information about the spatial origins through their links back to V1 but they do not otherwise make the locations explicitly available.

At a higher level in the ventral pathway, nodes in a recognition network representing object types for familiar entities are activated to different degrees, depending on how many of their features are present but regardless of how these are conjoined¹. The recognition network is a long-term store of knowledge used to identify objects but not to represent the current scene. In this preattentive or inattentive phase, features activate any object types with which they are individually consistent, and may inhibit those with which they conflict, activating particular recognition nodes to differing degrees depending on the level of feature support. Topdown expectations may also prime predicted objects. Because the feature access is nonselective, illusory conjunctions will be activated as well as correct conjunctions. Figure 1 shows the nodes for a red cross, and, through associative priming, for hospital, being activated by what is actually a yellow cross and a red heart. The initial access to identities on the first pass is very fast, taking only 150 ms or so (Thorpe, Fize and Marlot 1996), although not all are later confirmed

¹ Note however that some physical conjunctions may create emergent features, which could also play a part in activating only recognition units with which they are compatible (see, for example, Treisman and Paterson, 1984).

and consciously seen. So far the information is mostly unconscious and not explicitly available, although it may trigger responses based on the semantic categories activated (see below).

Figure 2 about here

Our hypothesis (Treisman and Gelade, 1980) was that incorrect conjunctions are weeded out by focused attention. Figure 2 shows the attention window narrowly focused onto one object location and suppressing any features outside the currently selected area. One way this might be implemented is through a reentry process, controlled by parietal areas, acting back on the map of locations, perhaps in V1, to serially select the contents of different locations (Treisman 1996). An “object file”, addressed by its location at a particular time, is formed to represent the attended object (Kahneman, Treisman and Gibbs, 1992). Within it, the selected features are bound and then compared to the set of currently active representations in the recognition network to retrieve the identity of the object – in this example the moon. Conscious perception depends on these object files².

The window of attention set by the parietal scan can take on different apertures, to encompass anything from a finely localized object to a global view of the surrounding scene. The resulting object file may represent the scene as a whole (e.g. an ocean beach), a pair of objects within the scene (e.g. a woman walking her dog), or even a single part of one object (e.g. the handle of a cup). In combination, these samples at differing scales build up a representation both of a background setting and of some objects within it.

² While conscious perception in the theory depends on object files, the converse may not be the case. Binding can sometimes occur without the bound objects becoming consciously accessible. For example in the negative priming paradigm with only two objects present, accurate binding can happen without attention. The attended object is consciously and correctly bound, and the remaining features belonging to the unattended object are bound by default, with the resulting object tokens surviving sometimes for several days or weeks (Treisman and DeSchepper, 1996). However, the unattended shape may fail to be consciously registered.

This account provides a dual representation of object identities, one implicit and unconscious and the other explicit and conscious: On the one hand, there is a continuum of activation in the recognition network of object types, modulated both by expectations, and by the amount of feature support allowed through at any time by the window of attention. In addition, the attended features are assembled to form a few explicit object tokens or “object files” that encode information from particular objects in their particular current instantiation, specifying the spatial relations and conjunctions of features, and mediating conscious experience. The number of object files that can be set up and maintained in working memory is very limited – only two to four bound objects (Luck and Vogel, 1997). However the semantic knowledge primed in the recognition network may be much richer (see discussion below).

Evidence for feature integration theory (FIT)

The evidence that led to this theory includes the following findings:

1. Search for a conjunction target requires focused attention whereas targets coded by separate features can be found through parallel search. This early claim was supported by several experiments (Treisman and Gelade 1980), but its generality was soon challenged. When the features of the conjunction are defined by separate dimensions and are highly discriminable, a strategy for guiding the search can bypass the need to bind every distractor (Treisman 1988; Wolfe, Cave and Franzel 1989; Treisman and Sato 1990). Because the strategy also depends on spatial selection (controlled from the relevant feature maps), this strategy fails for conjunctions of parts within objects, for example a rotated T among rotated Ls, where both parts are present in every item and only their arrangement changes. It also cannot be used when the target is unknown and defined as the odd one out in a background of known conjunctions (Treisman and Sato, 1990).
2. Participants make binding errors if attention is diverted or overloaded. When we flashed displays of three colored letters and asked observers to attend primarily to two flanking

digits, they reported many illusory conjunctions such as a red E or a green T when the stimuli were actually a red O, a blue T and a green E (Treisman and Schmidt 1982).

3. A spatial cue in advance of the display helps the identification of a conjunction target much more than the identification of a feature target (Treisman 1988)
4. Boundaries or shapes defined only by conjunctions are difficult to see, whereas feature-defined boundaries give easy segregation and shape recognition (Treisman and Gelade 1980, Wolfe & Bennett, 1997)). It seems that we cannot “hold onto” more than one conjunction at a time.
5. Priming can be object-specific. For example, a letter primes report of a matching letter when it reappears in the same frame, even after the frame has moved to a new location (Kahneman, Treisman et al. 1992). This suggests that temporary episodic object tokens are set up to represent objects in their current specific instantiations, and to bind their features across space and time.
6. Search for conjunctions activates the same parietal areas as shifting attention spatially (Corbetta, Shulman, Miezin and Petersen, 1995).
7. Patients with bilateral parietal lesions lose their ability to localize objects spatially. Since this rules out spatial attention, FIT predicts that they should experience problems with binding and with individuating objects. This prediction is confirmed (Robertson, Treisman, Friedman-Hill and Grabowecky, 1997; Humphreys, Cinel, Wolfe, Olson and Klempen, 2000). The patient we studied saw illusory conjunctions even with long displays of only two objects. He was unable to find conjunction targets in visual search tasks but had no problem searching for targets defined by a single feature. If the deficit reflects a perceptual space that has collapsed down to a single functional location, it could also explain the simultanagnosia. Our account of these deficits is depicted in Figure 3. The model gives a way of integrating the two symptoms described by (Balint 1909) - localization failures and simultanagnosia - while adding the third prediction, that these patients should have major problems with binding.

Binding, Perceptual Organization, and Visual Search

The first new experiment that I will describe, done in collaboration with my students, Evans and Hu, explored the effect of perceptual organization on search for conjunction targets, specifically, what determines the units that are entered into object files. A debate over whether attention selects locations or objects has been going on for years. Debates that last that long usually get the answer “both”. It is certainly the case that one can attend to a particular currently empty location and that this causes a faster response to whatever appears in that location. But it is also true that object characteristics can mould attention to fit them, and that responses are triggered faster to two parts of the same object than to equidistant parts of two different objects (Egley, Driver and Rafal, 1994).

But what defines an object before it has been selected and attended? In feature integration theory, preattentive grouping must take place either in the separate feature maps, where it will be defined by separate colors, orientations, directions of movement and so on, but not by conjunctions of those features, or in the map of filled locations where properties such as connectedness, proximity, and colinearity could play a role in defining the candidate objects for focused attention. Grouping can also be influenced by top-down control, in a recursive interaction.

Balints' patients may also provide some useful empirical evidence on what counts as a perceptual unit. Their simultanagnosia allows them to perceive only a single object, so an object may be whatever they can see at any given time. The circularity can be broken by seeing how well their responses predict attentional performance in normal participants. Humphreys and Riddoch (1993) showed a Balints' patient an array of dots in either one or two colors. He was unable to tell whether the array contained two colors or not, since he saw only one dot of the

many presented. However when heterogeneous pairs of dots were joined with connecting lines he was suddenly able to do the task. The lines made each pair into a unitary object so that two dots of different colors were now consciously accessible instead of one. However, when the lines connected matching pairs he failed. Robertson and I tried replicating this experiment with our Balints' patient. Unfortunately he did not show the effect. However he did make an intriguing remark that is also relevant to our question. He said "You know I can only see 1 or 2 of the whole bunch of them". Now where was that "bunch" that he knew about but allegedly could not see? Our guess is that he occasionally opened a larger object file, letting in a patch of dots as a single object. I will return to that observation later in the chapter.

Inspired by Humphrey's patient, Evans, Hu, and I explored a prediction about perceptual units and binding in normal participants. The target in our visual search task was a conjunction of shape and color – a blue circle among green circles and blue triangles. We used displays with lines that either connected the elements or left them unconnected. The lines connected either mixed or matching pairs (see Figure 4). The Balint's data suggest that connecting lines make perceptual units out of the pairs. If so, the linking might help when the pairs match, since there would be fewer items to search. But it should hinder when they are mixed because each object would contain both target features – circle and blue. Attention would have to narrow down to the separate shapes within the pairs, breaking the natural units. Search should be slower, and the risk of illusory conjunctions should increase.

The four conditions, Mixed Linked, Mixed Separate, Matching Linked, and Matching Separate, were randomly mixed within blocks. There were three display sizes: for each condition: 4 pairs in a 4x3 matrix; 9 pairs in a 5x5 matrix and 16 pairs in a 7x7 matrix. The 3x4 and 5x5 matrixes were randomly located in the 7x7 matrix, which was centered on the screen, so the densities and the average distances of elements from the initial fixation point were

approximately matched. A target was present on half the trials. The displays remained visible until participants responded. There were 12 participants, all students at Princeton University.

Participants missed far more targets in some conditions than in others, implying that they sometimes stopped the search without completing it, making it difficult to compare search times across conditions. To correct for the different numbers of missed targets, we reasoned as follows: If participants missed targets with probability p in displays of n items, they presumably searched only $(1-p)$ of the n items. We calculated the search rates replacing the objective display sizes with these estimates of the number of items searched. Since the functions appeared to increase linearly with the number of items and the slope ratios of target absent to target present were close to 2 to 1 (averaging 2.1), we assumed that they reflected serial self-terminating search. To get a concise summary of performance in the four conditions, we took the mean of the slopes for target absent and twice the slopes for target present (since only half the items would be checked in a self-terminating search). These are shown in Table 1.

Table 1 about here

The Mixed Linked condition, as predicted, gave by far the worst performance – search times per item nearly double those for the Matching Linked, and much worse than those for the Mixed Separate. The object files that preattentive processes deliver up for attentive checking in this condition are poorly adapted to the task of finding the blue circle target, because each pair creates a potential illusory conjunction. It has to be broken up perceptually before being rejected.

Components of Binding

Once the candidate objects for attention are specified, how do we actually bind their features? The second experiment, done with Seiffert and Weber, is an attempt to analyze the binding process and to see what brain mechanisms are involved. We distinguish three different components to the binding task in visual search: shift, suppress, and bind.

1. We must shift the attention window in space to select one object after another.
2. We must exclude or suppress features of other objects, to prevent illusory conjunctions
3. Finally we must bind the selected features together

Figure 5 about here

We used a difficult conjunction search task, initially tested by Wolfe, & Bennett, (1997), to try to separate the contribution of each of these components. The target was a plus with a blue vertical and a green horizontal bar. It appeared among either conjunction distractors – pluses with a green vertical and blue horizontal bar, or feature distractors with a unique feature not shared by the target (see Figure 5). In one experiment, the feature distractors differed from the target either in color or in orientation. They were a plus tilted 20 degrees left or right.. or a plus with a purple bar. In another experiment, they differed from the target either in color or in shape. They were a plus with a purple bar or a T with blue vertical and green horizontal bars. When the distractors have a purple bar or tilt, or when they differ from the target in a simple feature of shape, there should be no need to bind. They can be rejected on the basis of the unique nontarget feature. But when the target and distractors have the same colors and bars, the only way to determine whether they match the target is to bind the color, orientation or T feature to the bars for each distractor in turn.

Figure 6 about here

We presented strings of sequential displays, consisting of two components: 8 items in successive cued locations that could contain the target with 50% probability (we call this the target sequence), and a fixed set of five additional distractor items in other randomly selected uncued locations (we call this the surround). At the end of each string of 8 displays, participants were asked whether a target (a plus with a blue vertical and green horizontal bar) was present anywhere in the sequence. It could only appear in one of the cued locations. We isolated the binding component by comparing detection rates when the target sequence consisted of conjunction distractors and when it consisted of feature distractors. We tested the spatial

shifting component by presenting the items in the target sequence in different randomly selected locations, each one precued by a black dot, and comparing this shifting sequence with a sequence that appeared in a single cued location. We tested the distractor suppression component by varying the nature of the items in the surround. The five additional distractors could be conjunction distractors or feature distractors. The conjunction distractors should need active suppression to prevent their features from forming illusory conjunctions with those in the cued location.

Figure 7 about here

Figure 7 shows the error rates (false alarms plus omissions) in each condition. There was a large effect of shifting attention around in the display (significant, $F(1, 8) = 61.28, p < .001$). There was also a large effect of conjunction versus feature distractors in the sequence of cued locations ($F(1,8) = 37.86, p < .001$). This is the component that reflects the need to bind the colors and bars. There was also an interaction between the effects of shifting and binding ($F(1, 8) = 42.8, p < .001$). When the relevant location was fixed, there was little effect of the cued distractors. They interfere mostly when attention must be reset for each item. Finally there was no overall effect of having conjunction versus feature distractors in the surround. Looking in more detail at the data for the Shifting sequence only, we found that there was a significant effect contrasting conjunction with color distractors in the surround (Figure 8a), and there was also an effect in the second experiment contrasting conjunction with color or with shape distractors (see Figure 8b). Thus there may be something special that makes orientation feature distractors in the surround harder to suppress than color or shape feature distractors and as hard as conjunction distractors. Finally in both experiments, when there were conjunction distractors in the cued sequence there was no additional decrement from replacing feature distractors in the surround with conjunction distractors. A possible reason could be the greater heterogeneity in the display with two different distractor types (feature and conjunction) compared to one (all conjunction distractors). We are currently testing with two kinds of feature

distractors (e.g. color distractors in the cued locations and orientation distractors in the surround), and we find that, with heterogeneity controlled in this way, there is clearly more interference when conjunction distractors are present.

Figure 8 about here

Brain imaging offers a new tool for dissecting separate operations. We were interested in testing the reality of these separate components of binding by relating them to different brain areas as well as to separate behavioral effects (Seiffert, Weber, and Treisman, 2003). We tested 8 participants doing the first experiment with Color and Orientation as feature distractors in the Princeton Siemens fMRI brain scanner. To increase the sensitivity in the Shifting attention conditions, we tested only one of the Fixed Sequence conditions (conjunction distractors in both Sequence and Surround). Previous studies (eg Corbetta, Shulman et al., 1995) have shown parietal activity in conjunction search, but most did not distinguish shifting from suppression or from binding. Comparing the activation in the corresponding Shifting and Fixed Sequence conditions, we found the expected parietal activation, and also plenty of occipital activation, probably due to the presence of motion. More interesting, we found two separate areas that were activated selectively in the binding comparison – Conjunction versus Feature distractors in the cued sequence. One was a medial frontal area, close to or in the anterior cingulate. This tends to be activated when there is some form of conflict present. In binding, there may be competition between relevant and irrelevant features to be integrated into the currently selected object. The other area showing activity selective to binding was the left anterior insula. We also found this area in another conjunction search experiment, so we think it may be a real effect. But its function is more mysterious. The anterior insula appears to be activated in a wide variety of rather specific but diverse situations, ranging from disgust (Critchley, Wiens, Rotshtein, Öhman and Dolan, 2004), to imitation of facial expressions (Carr, Iacoboni, Dubeau, Mazziotta and Lenzi, 2003), to articulation in speech (Dronkers, 1996; Wise, Greene, Buchel and Scott, 1999), as well as to several other equally disparate situations. It seems unlikely that what they

all have in common is binding! Of course there may well be different specialized sub-areas within the insula involved in these different tasks, one of which may be concerned with binding. Further research will be needed to explore that possibility.

There is no complete story yet on the neural instantiation of binding. My speculation is that it reflects a process of re-entry after the first initial pass through the sensory hierarchy, returning to the early visual areas where fine localization and detailed features can be retrieved. The idea of re-entry, (first proposed in the context of memory by Damasio (1989), is very much in the air these days. For example it is used in the Reverse Hierarchy Theory proposed by Hochstein and Ahissar (2002) and in the model of visual masking by object substitution proposed by Di Lollo, Enns and Rensink (2000). I suggested that it may also play a role in binding, as follows (Treisman, 1996). The first response to visual stimulation may activate feature detectors in early striate and extra-striate areas that connect automatically to the temporal lobe object nodes with which they are compatible, and perhaps inhibit those with which they conflict. To check whether the conjunctions are real, the features must be retraced to the early visual areas V1 or V2 where localization is more precise. Parietal areas may then control a serial reentry scan through these areas to retrieve the features present in each. The binding itself might involve the area we found in the anterior insula combining the selected features to form an integrated object representation.

Distributed Attention

The second part of this paper explores the effects of deploying attention more globally over several objects at once or over the display as a whole. The hypothesis is that this provides different information from that available in the focused binding mode. Figure 9 shows the attention window encompassing three objects instead of one, making all their features available to potential object files. A separate choice can then be made of whether to open separate object

files for the individual elements, running the risk that a random selection of features may result in illusory conjunctions. The yellow cube in the figure is an illusory conjunction of two attended features that actually belong to different objects. Because it is in an object file, it could become consciously available. Alternatively one could choose to open a single object file for the attended group as a whole, making a different set of properties available, including the global shape, global boundaries, and global relations between elements. The “whole bunch of dots” reported by our Balint’s patient may reflect processing of a global patch as a unitary object.

Figure 9 and Table 2 about here

What useful information might this global deployment provide? Table 2 summarizes the following suggestions.

1. First, it could be an efficient way of revealing the presence of a target feature anywhere in the display. If we are looking for something “red” and red appears in the global object file, we can detect its presence, and then rapidly home in on it in the location map with more narrowly focused attention.
2. It might define global boundaries between groups of objects where the groups differ in some feature that is shared by elements within each group and differs for elements between groups. Groups defined only by conjunctions of features do not create boundaries that are perceptually available. (Treisman and Gelade, 1980).
3. It should give the global shape of the objects, at some expense to the local objects that compose them (Navon, 1977).
4. It should provide some general parameters like the global illumination of the elements, allowing us to detect an odd object that is lit from a different direction or oriented differently in three dimensional space, even though the unique object seems to be distinguished only by a conjunction of shape and contrast (Enns and Rensink, 1996)

5. With more complex natural scenes, global attention may also give us the gist of the scene – for example a mountain landscape, or a kitchen, perhaps through priming from disjunctive sets of features (see below).

6. Finally our current hypothesis is that it may offer some simple statistical properties of sets of similar objects.

Statistical Processing with Distributed Attention

The world is full of sets that vary on a range of different dimensions, for example the cars in a parking lot, or the apples on a tree. It would be wasteful and probably beyond our attentional capacity to specify each element in the representation we form. Instead we may generate a statistical description, including the frequencies of different element types, the mean, the range and the variance of sizes, colors, orientations that they encompass, facilitating rapid decisions based, for example, on the quality of the fruit, the density of the traffic, or the rockiness of the terrain (see Figure 10).

One important proviso is that this statistical information should be restricted to features, since it comes from pooled feature maps rather than from individuated objects. I recently tested this prediction by asking participants to judge the frequencies either of features or of conjunctions. I presented brief displays of colored shapes - the letters, O, X, and T in red, green and blue - and asked people to estimate proportions, for example what proportion were green, or were Ts, or were green Ts. I post-cued which feature or conjunction was relevant (Figure 11). On each trial, participants saw the display for 500 ms, then 500 ms blank screen, and then a single color patch or a single white letter or a single colored letter as a probe. The task was to say what proportion of the display that particular feature or conjunction represented. Participants were good at judging the proportions of the separate features that were present, but very poor at judging the proportions of conjunctions. For example, they gave similar

estimates for the proportion of blue Xs in the 2 displays of Figure 12 a and b. The two displays contain exactly the same proportions of the different features – half blue letters, half Xs, a quarter each red letters, green letters, Ts, and Os, but the features are differently bound in the two conditions. The first display actually has 33% blue Xs and the second has none. Yet the mean estimates hardly differed: They were 15% and 11%. Global attention gives us good estimates of feature proportions, but it does not yield proportions for bound objects.

Figures 11 and 12 about here

Also exploring this framework of statistical perception, my student Chong and I have done some research on judgments of mean size with heterogeneous arrays of circles. Our starting point was a paper by (Ariely 2001) in which he showed that the mean size of a set of circles is recognized better than any individual size in the set, and that extraction of the mean is not impaired by increasing the number of items in the display. We found this result quite surprising and it piqued our curiosity. First we compared the threshold for judging mean sizes with the threshold for judging the size of an individual circle presented alone, to see just how good the statistical judgment is (Chong and Treisman, 2003). With a single item, we can focus attention and give it our full resources, so we might expect to get much more accurate discriminations. We used forced choice judgments of which side of a display had the larger size, or the larger mean size, and we added a third condition in which people judged the sizes of circles in two homogeneous sets (Figure 13). We found very little difference between the three conditions. Computing the mean size in heterogeneous arrays was about as accurate as judging the presented size of single circles or of homogeneous arrays. Next we tried varying the exposure duration. If the mean is computed by serially adding each individual size and dividing by N, we would expect a clear deterioration as the available time was reduced. But we found very little. Moreover, the decrement was no larger for the mean than for the homogeneous displays (Figure 14). The measure is the size threshold, expressed as a percentage of the difference in diameter required to give 75% correct discrimination.

Figures 13 and 14 about here

Do the displays have to be simultaneously available, or can we hold onto one in memory? We compared simultaneous with successive presentations of the two displays (Figure 15), and used two delays in the successive condition (100ms and 2 seconds) to see whether performance depended on perceptual availability or whether the first display could be briefly retained in visual memory. Again there was surprisingly little difference between any of these conditions (Figure 16). The mean judgment showed a small decrement with the two-second ISI but so did the single circles. Either we compute the mean and hold that in memory, or we store the display as such and make the comparison when the second display appears.

Figures 15 and 16 about here

Were participants really computing the mean? Or might they be, for example, looking just at the largest size. To find out, we tested them on displays with 12 items to the left and 12 to the right of the center (see Figure 17). The task was to say which side of the display had the larger mean size. The two sets could either be drawn from the same distribution (one of the following: Uniform (four equiprobable sizes), Two Peaks (two equiprobable sizes – the largest and smallest of the four previous sizes), Normal (an approximation to a normal distribution), and Homogeneous (a homogeneous display of circles all at the mean size of the other displays), or they could be generated from any possible pair of these different distributions. Again people were surprisingly good at making these statistical judgments (Figure 18). They were only slightly worse comparing across two different distributions than comparing two samples from the same distributions. See, for example, the threshold for Normal versus Normal and for Normal versus Homogeneous – they are about the same, although the sizes are identical in the Normal versus Normal and have no overlap in the Normal versus Homogeneous. This suggests that participants really were computing the mean.

Figure 18 about here

Our hypothesis is that statistical processing occurs automatically when attention is globally distributed over a set of similar items. To test this we tried a couple of dual task paradigms, both using the same displays of 12 circles of four different sizes (Chong and Treisman, 2004). We combined tests of statistical judgments of the mean size with one of two visual search tasks, one that seems to require focused attention to each item in turn and one that gives popout or parallel processing, and that, according to our hypothesis, is done with global attention to the display as a whole. The focused attention task was search for a closed circle among circles with gaps, and the parallel processing version was search for a circle with a gap among closed circles (Figure 19). Participants saw the display and did the search task as quickly as they could. As soon as they pressed a key to signal target present or absent, two probe circles appeared and participants made a forced choice judgment of which matched either the mean size or the size of an individual circle sampled at random from the display. The location of the relevant circle was post-cued by a small location marker presented at the same time as the two probe circles. We obtained the predicted interaction (significant, $F_{(1, 11)} = 9.6, p < .05$). : Performance was better on the mean judgment (threshold 22%) when the concurrent task was the popout task using global attention, compared to when the concurrent task required serial search (threshold 25%), even though the display duration (determined by the search reaction time) was considerably longer with the focused attention task. On the other hand participants were better on the individual item size judgment when the concurrent task required focused attention (threshold 24%) compared to when the target popped out (threshold 28%). .

Figure 19 about here

Does the extra task impair the statistical processing at all, or is the mean size computed automatically? We compared performance in the dual task with attention globally deployed and performance when the statistical task was the only one required. There was no decrement with the global attention version of the dual task, relative to single task thresholds. It seems that when attention is spread over the display as a whole, the mean size is automatically extracted

with no demand for additional resources.

We tested the automaticity idea in a different way as well, by seeing whether the extraction of the mean size is automatic in the sense of producing implicit priming. We presented an array of 12 circles of just two sizes, large and small, followed by a speeded same-different matching task on two new circles (Figure 20). One or both the target circles would match either the mean of the previous set, or one of the two presented sizes, or some other size, either within the range of presented sizes or outside the range. If observers automatically form a representation of the mean size when they see a heterogeneous display, we might expect priming for subsequent perception of an individual circle at the mean size, and that is what we found (Figure 21). The latency to respond same or different was shorter for circles primed by the mean size than for any other size, including the ones that were physically presented.

Figures 20 and 21 about here

Finally Chong and I tested the psychological reality of the mean size by seeing whether it would generate false positive responses in a visual search task. Would people see illusory targets when the real target matched the mean size of the search display? We controlled for discriminability by matching the difference between the target and the nearest distractor size. The stimuli were selected from 6 possible sizes, two distractor sizes that were always present (fixed), a third optional distractor size that was either the largest or the smallest size (extreme), and another size that was either the target size or another distractor size (critical). The choice of which extreme distractor size was included determined which of the critical sizes matched the mean size of the display (see Figure 22). The displays contained 12 circles, including 3 each of the two fixed distractor sizes, 3 of one extreme size, and 3 of one of the critical sizes. By adding either the large or the small extreme distractor size, we could shift the mean size of the display to one or other of the two possible critical sizes (see Figure 22). There were four conditions: Target present with targets at the mean size, target present with targets at the other critical size,

target absent with additional distractors at the mean size, and target absent with additional distractors at the other critical size. Note that each of the two critical sizes played all four roles across different conditions. All that differed was the size of the extreme set of distractors. The conditions were randomly mixed across trials and participants were shown the relevant target circle before each display. As predicted, participants made more false alarms on target absent trials when the target matched the mean size of the search display (mean 34% compared to 28% when the target was not the mean size, $t_{(13)} = 2.437, p < .05$), suggesting that seeing a set of similar stimuli does indeed automatically generate a mental representation of the mean size, (and perhaps of other statistical parameters as well).

Figure 22 about here

So what conclusions emerge from these experiments on statistical processing? Discrimination of the mean size of a set of objects seems to occur very rapidly and about as accurately as discrimination of the size of a single individual object. It seems to be automatic, provided that attention is globally deployed. Even when the display contains just two sizes, a representation is formed of the mean size, which primes subsequent perception more strongly than do either of the sizes that were actually presented. We plan to look at other dimensions and other statistical measures, but we think we have some evidence for a separate mode of processing visual information using distributed or global attention, which contrasts with the focused mode used in binding and individuating separate objects.

Perception of Meaning and Gist

Returning again to the general framework: Is this enough to get us seeing the everyday world? I am not sure. There are some puzzling results in the literature. So far I have discussed results obtained with simple displays, designed to tease apart particular mechanisms underlying performance. However, the world we deal with is normally rich, complex, full of meaning and emotional associations. I will contrast some research findings that deal with this richer world and

ask what implications they have for a more complete understanding. One recent paradigm studying change blindness fits well, suggesting that we deal with the complexity quite simply by ignoring it. Rensink, Simons and others showed that, if simple low level cues to change (like apparent motion and visual transients) are removed or masked, we are very slow to detect quite salient changes in natural scenes (O'Regan, Rensink and Clark 1999; Rensink, 1998; 2000; Simons 1996; Simons and Levin 1998). If we happen to attend to the changing object, we are likely to become aware of the change, but otherwise it may take us several seconds to find it. O'Regan et al. suggest that we briefly summarize the scene and rely on the world itself to store the information in case we need to refer to any part of it in detail. Note, however, that the change detection task requires us not only to form representations of all the objects in the scene, but also to store and compare them across successive frames. It does not rule out more complex perception, which is instantly lost or erased by the next display (Wolfe 1998).

The other contrasting line of research probes immediate perception rather than memory and suggests a much larger capacity and the possibility of attention-free semantic processing. It goes back to Potter's demonstrations in the 70's (Potter 1975), in which she showed strings of natural scenes at high speed, successively in the same location, and asked people to find semantic targets, for example, a clown, or an airplane, or a donkey. She found that they could do this pretty well at presentation rates of up to 10 pictures/second, although memory for the nontarget pictures was very poor. More recently, Li, VanRullen, Koch and Perona (2002) have looked at the attention demands of this task and found a surprising result. They occupied attention at a foveal location by asking people to decide whether 5 small letters, - Ts and/ or Ls in random orientations - were all the same or contained one odd letter. This is a very demanding task and leaves people badly impaired in discriminating a single target letter in the periphery. However, when Li et al. briefly flashed a picture of a natural scene in the periphery, and asked people to determine whether it contained an animal or not, performance was just as good when

attention was focused on the foveal T/L task as when the scene classification task was the only one required.

So we have an apparently paradoxical pair of findings. You cannot even see a salient object like an airplane engine coming and going in the change blindness picture, but you can detect the presence of an unknown animal in a 50 ms exposure while focused on a completely different set of stimuli. Can we reconcile those findings? How would they fit the framework I proposed? The animal detection task clearly is not using the binding mode with focused attention, since that is devoted to the foveal letter task. But my suggestion was that even without attention, on the first rapid pass through the visual system, we can detect disjunctive sets of features. In searching for a target category, participants may be set to sense, in parallel, a highly overlearned vocabulary of features that characterize a particular semantic category. For example animals might be signaled by the presence of a compact curvilinear figure against the background, a set of legs, a head, an eye, wings or fur. If present, these animal features would prime the general, semantic category and could perhaps distinguish a slide with an animal from one without (cf Levin, Takarae, Miner and Keil 2001). Evans and I have shown that detecting an animal in RSVP strings of natural scenes is much harder when the scenes also contain humans than when they do not (58% vs 76% hits), presumably because shared features required observers to raise their criteria. On the other hand, detecting a vehicle is not affected by whether or not humans are present in the distractor slides. Vehicles and people share no obvious features. Participants were also quite bad at locating the target animal (to the left, center or right of the scene), suggesting a failure to bind the features to their locations. Of the 67% correct detections, only 53% were also correctly located, (where chance was 33%). The target animal was correctly identified on only 46%% of the hit trials (e.g. as a bear, a butterfly, or a goldfish), although observers had at least some information about the superordinate category

(e.g. mammal, insect, or fish), and were correct on the superordinate category on 81% of trials. The categories should be available from one or more distinctive features (wings, fur, etc.).

One objection that might be raised is that these animal features are physically more complex than the colors, shapes, and orientations used in my earlier research. They do not sound like features in the usual sense of the term. Firstly, however, I do not claim in the theory to have defined *a priori* any specific set of features. They cannot be specified independently of the contrast set from which they are to be discriminated in any given setting. The set of feature detectors used in any task is an empirical question, to be tested with a set of converging operations such as those shown in Table 3. This is a boot-strapping account, but if the tests do converge, we have learned something. Secondly, the animal features may not be more complex for the visual system than the “simple” colors, line orientations, directions of motion and so on tested in the usual psychophysical experiments. We have evolved to detect the natural components of real world objects and their natural textures and colors. Gross, Rocha-Miranda and Bender (1972) showed that single units in area IT respond to monkey hands or faces and Tanaka (1993) found many units in IT that respond to what look like elementary components of natural objects.

Table 3 about here

Animals are a natural category with which we have evolved, so it might make sense to have fur and eye detectors. However another possible objection arises when we consider that Li et al. got similar results using the category of vehicles. Admittedly, it is harder to tell an evolutionary story about the features of manmade objects like cars and airplanes. I would not expect wheel and engine detectors to be part of our innate repertoire. However, there is a great deal of plasticity in the brain, and feature detectors may be created to carry out any discrimination that proves useful. Evolution would provide the plasticity at the feature level to enable whatever elements prove consistently useful to be detected automatically. Freedman and his colleagues

(Freedman, Riesenhuber, Poggio and Miller 2001; 2002) recorded from single units in prefrontal cortex of monkeys, discriminating members of the cat family from various breeds of dogs, together with many intermediate morphs. When the reward contingencies were changed from distinguishing cats and dogs to distinguishing creatures that varied on the orthogonal dimension – dog/catness versus cat/dogness, - the same units learned to respond selectively to one of the new categories.

Figure 23 about here

Given the possibility of flexible feature sets, we may be able to account for visual category detection within the framework I am using, as shown in Figure 23. Once the category is primed, if the target animal or vehicle is still present and attention can home in on it, we can form an object file to make it consciously accessible. If not, we may be able to tell simply from its features which animal it is likely to be – cats seldom have wings and beaks. But we might make mistakes.

Why would this framework not also predict easy change detection? There are a number of differences which could make change detection harder than category detection: First, change detection depends on comparing successive displays, thus involving both visual memory and the need to match a current display to a stored one; secondly participants are not cued in advance with the relevant category (or the relevant sets of features) to look for. Primed detection of a prespecified category could potentially be mediated by unbound features, whereas the change detection task imposes a much heavier load involving both binding, non-selective memory, and a non-selective comparison process.

This paper has specified three ways of deploying attention and three types of information that they may provide: (1) Focused attention, offering conscious access to individuated objects with bound features in their current locations, viewpoints and configurations; (2) at the other end

of what is actually a continuum, global or distributed attention, offering global and statistical properties of groups of objects or the general layout of a scene; (3) rapid preattention, or more prolonged inattention, offering parallel and automatic access to sets of unbound features which can also prime semantic categories in a recognition network. Perhaps in combination, these parallel feature detectors, flexibly evolved through learning, together with the attention-controlled construction of a small set of explicit object files, and the statistical descriptions of sets of similar elements, may approach the outlines of an explanation for our amazing ability to see and interact with the world around us, even with the limited resources set by visual attention.

Acknowledgements

The research was supported by grants from the NIH Conte Center # P50 MH62196, from the Israeli Binational Science Foundation, # 1000274, and from NIH grant 2004 2RO1 MH 058383-04A1 Visual coding and the deployment of attention.

References

- Ariely, D. (2001). Seeing Sets: Representation by statistical properties. Psychological Science **12**: 157-162.
- Balint, R. (1909). Seelenlahmung des 'Schauens', optische Ataxie, raumliche Störung der Aufmerksamkeit. Monatschrift für Psychiatrie und Neurologie. **25**: 5-81.
- Carr, L., Iacoboni, M., Dubeau, M. J., Mazziotta, J. and Lenzi, G. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. Proceedings of the National Academy of Sciences **100**: 5497-5502.
- Chong, S. and Treisman, A. (2003). Representation of statistical properties. Vision Research **43**: 393-404.
- Chong, S.-C. and Treisman, A. (2004). Attentional spread in the statistical processing of visual displays. Perception and Psychophysics, in press.
- Corbetta, M., Shulman, G. L., Miezin, F. M. and Petersen, S. E. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. Science **270**: 802-805.
- Critchley, H., Wiens, S., Rotshtein, P., Öhman, A. and Dolan, R. (2004). Neural systems supporting interoceptive awareness. Nature Neuroscience **7**(2): 189-195.
- Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. Cognition **33**: 25-62.
- Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. Nature **384**: 159-161.
- Egely, R., Driver, J. and Rafal, R. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. Journal of Experimental Psychology: General **123**(2): 161-177.
- Enns, J. T. and Rensink, R. A. (1996). An object completion process in early vision. Visual search III. Gale, A. London, Taylor & Francis.

- Farah, M. J., Drain, H. M. and Tanaka, J. W. (1995). What causes the face inversion effect? Journal of Experimental Psychology: Human Perception and Performance **21**: 628-634.
- Freedman, D., Riesenhuber, M., Poggio, T. and Miller, E. (2002). Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. Journal of Neurophysiology **88**(2): 929-941.
- Freedman, D. J., Riesenhuber, M., Poggio, T. and Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science **291**: 312-316.
- Gross, C., Rocha-Miranda, C. E. and Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. Journal of Neurophysiology **35**: 96-111.
- Hochstein, S. and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. Neuron **36**: 791-804.
- Houck, M. R. and Hoffman, J. E. (1986). Conjunction of color and form without attention: Evidence from an orientation-contingent color aftereffect. Journal of Experimental Psychology: Human Perception and Performance **12**: 186-199.
- Humphreys, G. W., Cinel, C., Wolfe, J. M., Olson, A. and Klempen, N. (2000). Fractionating the binding process: neuropsychological evidence distinguishing binding of form from binding of surface features. Vision Research **40**(10-12): 1569-1596.
- Humphreys, G. W. and Riddoch, M. J. (1993). Interactions between object and space systems revealed through neuropsychology. Attention and Performance XIV. Kornblum, S. Cambridge, MA, MIT Press: 183-218.
- Kahneman, D., Treisman, A. and Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. Cognitive Psychology **24**: 175-219.
- Levin, D. T., Takarae, Miner, A. G. and Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. Perception and Psychophysics **63**: 676-697.

- Li, F., VanRullen, R., Koch, C. and Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. Proceedings of the National Academy of Sciences, USA **99**(14): 9596-9601.
- Luck, S. J. and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. Nature **390**: 279-281.
- McCollough, C. (1965). Color adaptation of edge-detectors in the human visual system. Science **149**: 1115-1116.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. Cognitive Psychology **9**: 353-383.
- O'Regan, J. K., Rensink, R. A. and Clark, J. J. (1999). Change-blindness as a result of "mudsplashes." Nature **398**: 34.
- Potter, M. (1975). Short-term conceptual memory for pictures. Science **187**: 965-966.
- Rensink, R. A. (1998). Mindsight: Visual sensing without seeing. ARVO abstract.
- Rensink, R. A. (2000). Visual Search for Change: A Probe into the Nature of Attentional Processing. Visual Cognition **7**: 345-376.
- Robertson, L., Treisman, A., Friedman-Hill, S. and Grabowecky, M. (1997). The interaction of spatial and object pathways: Evidence from Balint's syndrome. Journal of Cognitive Neuroscience **9**: 254-276.
- Seiffert, AE, Weber, M., and Treisman, A.. (2003). Isolating feature binding in visual search. Society for Neuroscience meeting in New Orleans (Nov. 8 to 12).
- Simons, D. and Levin, D. (1998). Change blindness. Trends in Cognitive Science **1**: 261-267.
- Simons, D. J. (1996). In sight, out of mind. Psychological Science **7**: 301-305.
- Tanaka, K. (1993). Neuronal mechanisms of object recognition. Science **262**: 685-688.
- Thorpe, S., Fize, D. and Marlot, C. (1996). Speed of processing in the human visual system. Nature **381**: 520-522.

- Treisman, A. (1986). Properties, parts and objects. Handbook of Perception and Human Performance. Thomas, J. New York, Wiley. **2**: 35, 1-70.
- Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. Quarterly Journal of Experimental Psychology **40A**: 201-237.
- Treisman, A. (1996). The binding problem. Current Opinion in Neurobiology **6**: 171-178.
- Treisman, A. and Gelade, G. (1980). A feature integration theory of attention. Cognitive Psychology **12**: 97-136.
- Treisman, A. and Sato, S. (1990). Conjunction search revisited. Journal of Experimental Psychology: Human Perception and Performance **16**: 459-478.
- Treisman, A. and Schmidt, H. (1982). Illusory conjunctions in the perception of objects. Cognitive Psychology **14**: 107-141.
- VanRullen, R., Reddy, L. and Koch, C. (2004). Visual search and dual-tasks reveal two distinct attentional resources. Journal of Cognitive Neuroscience **16**(1): 4-14.
- Wise, R., Greene, J., Buchel, C. and Scott, S. (1999). Brain regions involved in articulation. The Lancet **353**: 1057-.
- Wolfe, J. M. (1998). Inattentional amnesia. Fleeting memories. Coltheart, V. Cambridge, MA, MIT Press: 71-94.
- Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. Vision Research, **37**, 25-44.
- Wolfe, J. M., Cave, K. R. and Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human Perception and Performance **15**: 419-433.

Table 1**Search functions for the four conditions shown in Figure 4.**

(Slopes are means of 2(Target Present) and (Target Absent), calculated after correcting the display sizes for the proportion of targets missed on Target Present trials at each display size)

| | Mixed Pairs | | | Matching Pairs | | |
|-----------------|--------------------|-----------|-----------------------|-----------------------|-----------|-----------------------|
| | Slope Estimate | Intercept | Mean % Missed targets | Slope estimate | Intercept | Mean % Missed targets |
| Linked | 187 | 359 | 21 | 108 | 470 | 12 |
| Separate | 130 | 452 | 8 | 98 | 480 | 5 |

Table 2**Information Potentially Available with Distributed or Global Attention**

1. Target with unique feature (popout effect in search)
2. Global boundaries defined by features (not conjunctions)
3. Global shape
4. Global illumination and 3D orientation
5. Gist of the scene (e.g. mountain landscape, or kitchen)
6. Statistical properties

Table 3**Diagnostics for Separable Perceptual Features (see Treisman, 1986)**

Mediate parallel search

Support perceptual grouping and segregation

Can migrate independently in illusory conjunctions

Can be separately attended

Neural evidence - separate populations of single units

Show selective adaptation and selective masking

Figure Captions

Figure 1. Sketch of model before attention acts (preattentive processing) or when attention is focused elsewhere (inattention).

Figure 2. Sketch of model with attention focused on one object. The features in the selected location are collected in an object file where they are bound. Non-selected features are excluded from the object file but may still prime object nodes in recognition network.

Figure 3. Sketch of model for Balint's patient with bilateral parietal damage. The location map has been destroyed, preventing feature binding and the individuation of different simultaneously presented objects through spatial attention. As a result, binding errors are frequent and only one object file is available at any given time.

Figure 4. Stimuli used in visual search and grouping experiment (Evans, Hu, and Treisman, unpublished). The target was a blue circle (shown as speckled) among blue triangles and green circles (shown as grey).

Figure 5. Stimuli used by Seiffert, Weber and Treisman (unpublished) to test separate components of the binding process.

Figure 6. Example of trial in Components of Binding experiment. Each trial comprised a sequence of 8 cued items, interleaved with a spatial cue (a black dot) preceding each display. If present (50% trials), the target always appeared at a cued location. The distractors differed from the target in either a Feature (color or tilt), or the Conjunction of features. Distractors were added in the surround to test suppression of non-target items. The figure shows the first three displays in a trial in which the shift-attention, and suppress-surround components are needed,

but not the binding component for the cued sequence, since the cued items differ from the target in a simple feature (color).

Figure 7. % Errors and omissions made in the different conditions of experiment testing components of binding. Trials requiring shifts of attention are slower than those with fixed attention. When attention must be reset for each item, trials with conjunction distractors in the cued sequence are slower than those with feature distractors, suggesting that binding increases the difficulty. However there is no overall effect of conjunction distractors relative to feature distractors in the surround.

Figure 8. (a) Separating trials with color feature and trials with orientation feature distractors shows that tilt distractors are as difficult as conjunction distractors to suppress, but color feature distractors are easier.

(b). Separate analysis of shape and color distractors (in shifting attention trials only) shows that both feature types are easier than conjunction distractors to suppress. However, neither of the experiments shows increased difficulty when conjunction distractors in the surround are added to conjunction distractors in the cued sequence. This may be because display heterogeneity is greater with feature distractors in the surround and conjunction distractors in the sequence. Two different distractor types must be excluded rather than a single type.

Figure 9. Sketch of model with attention spread over three items rather than focused on one. The hypothesis is that this produces binding failures within the attention window.

Figure 10. Statistical properties may be automatically computed at the feature level when attention is distributed over sets of similar items.

Figure 11. Experiment testing participants' ability to estimate the proportions of different features or of different conjunctions. The set of relevant targets to be reported (e.g. red items, T's, or red T's) was specified immediately after the display was presented.

Figure 12. Examples of two displays with equal numbers of each feature type but different numbers of various conjunctions. Black represents blue letters. The display in (a) has 36% blue Xs and the display in (b) has none. The mean estimates were 15% and 11% respectively.

Figure 13. Examples of stimuli for size judgments. Participants decided which side had the larger size or the larger mean size.

Figure 14. Effects of exposure duration on mean thresholds for size judgment with single circles, homogeneous sets or mixed sets. The measure is the % diameter difference giving 75% accuracy in the forced choice judgment.

Figure 15. Procedure for comparing simultaneous presentation with successive presentation of two displays.

Figure 16. Effects of delay between the two sets of stimuli on thresholds for judging size or mean size.

Figure 17. Different distributions tested in mean size judgments. Comparisons were made both within and across distributions. H = Homogeneous, U = Uniform, T = Two peaks, N = Normal. The numbers indicate the mean diameter size in degrees of visual angle.

Figure 18. Mean thresholds within and across distributions. H = Homogeneous, U = Uniform, T = Two peaks, N = Normal

Figure 19. Design of experiment testing the effect of a concurrent task requiring either focused attention (search for a closed circle among circles with gaps) or distributed attention (search for a circle with a gap among closed circles) on two size judgment tasks – judgment of the mean size and judgment of the size of an individual post-cued circle.

Figure 20. Design of experiment testing whether perception of the mean size of the preceding display is favored in a subsequent same-different judgment task.

Figure 21. Mean response times as a function of which size from the preceding display was primed in the same-different judgment. The mean gives the shortest latencies, even compared to the sizes that were actually present in the display.

Figure 22. Schematic representation of stimuli to test the effect of target size in relation to the mean of the display in a visual search task. The Fixed distractor sizes appeared in all displays and one of the two Extreme distractor sizes was added. Depending on which was added, one or other of the Critical sizes became the mean size and the other the control size. One of these was designated the target size and could be either present or absent. When it was absent the other critical size was present as a fourth distractor size. Each display had three instances of each of the four relevant sizes. The target size was cued in advance of each display.

Figure 23. Sketch of model to account for detection of high-level perceptual categories. The suggestion is that sets of learned features may be detected in parallel and prime their category without being bound through focused attention.

Figure 1

First feedforward pass
(or later passes with attention focused elsewhere)

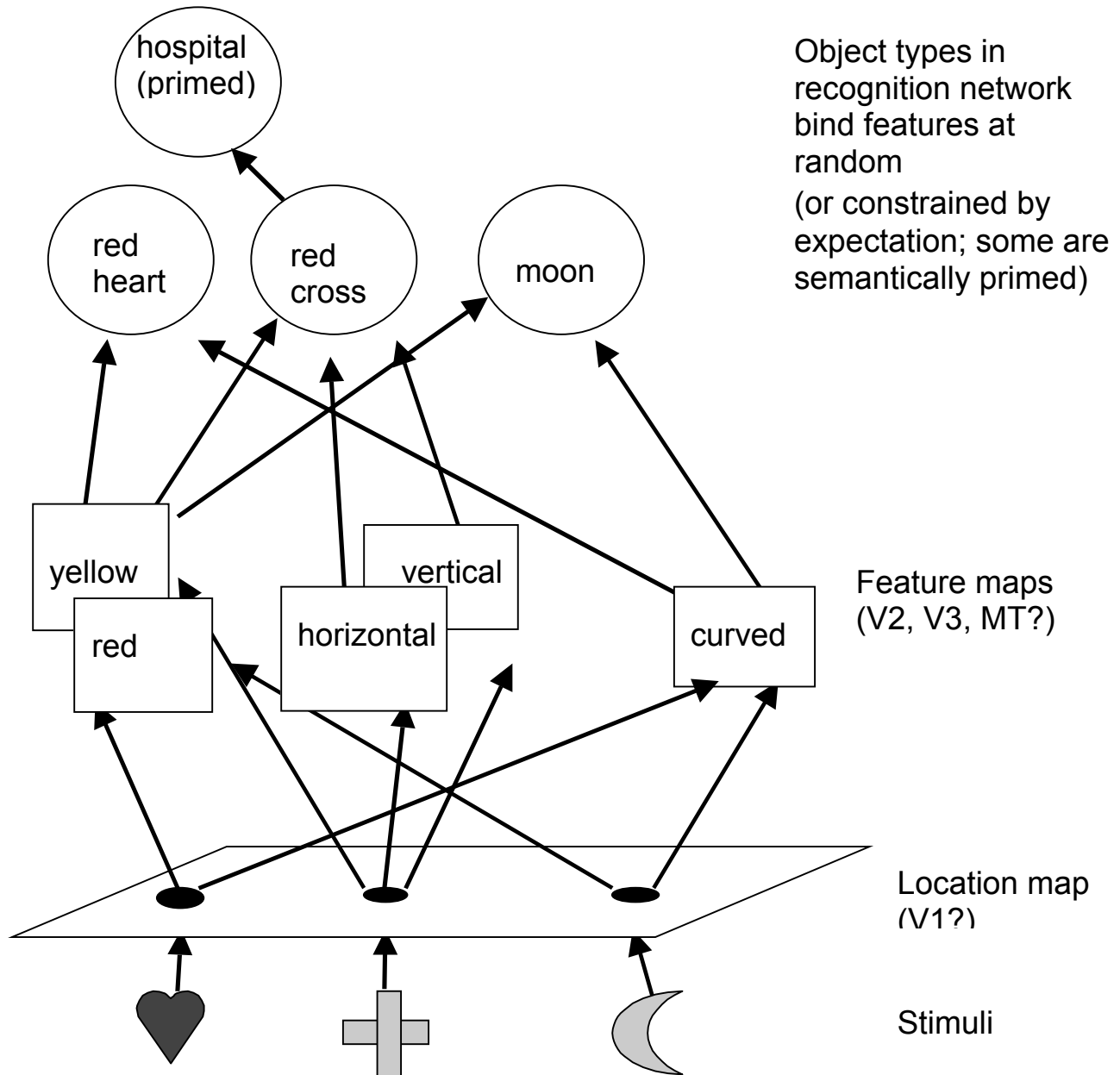


Figure 2

With focused attention

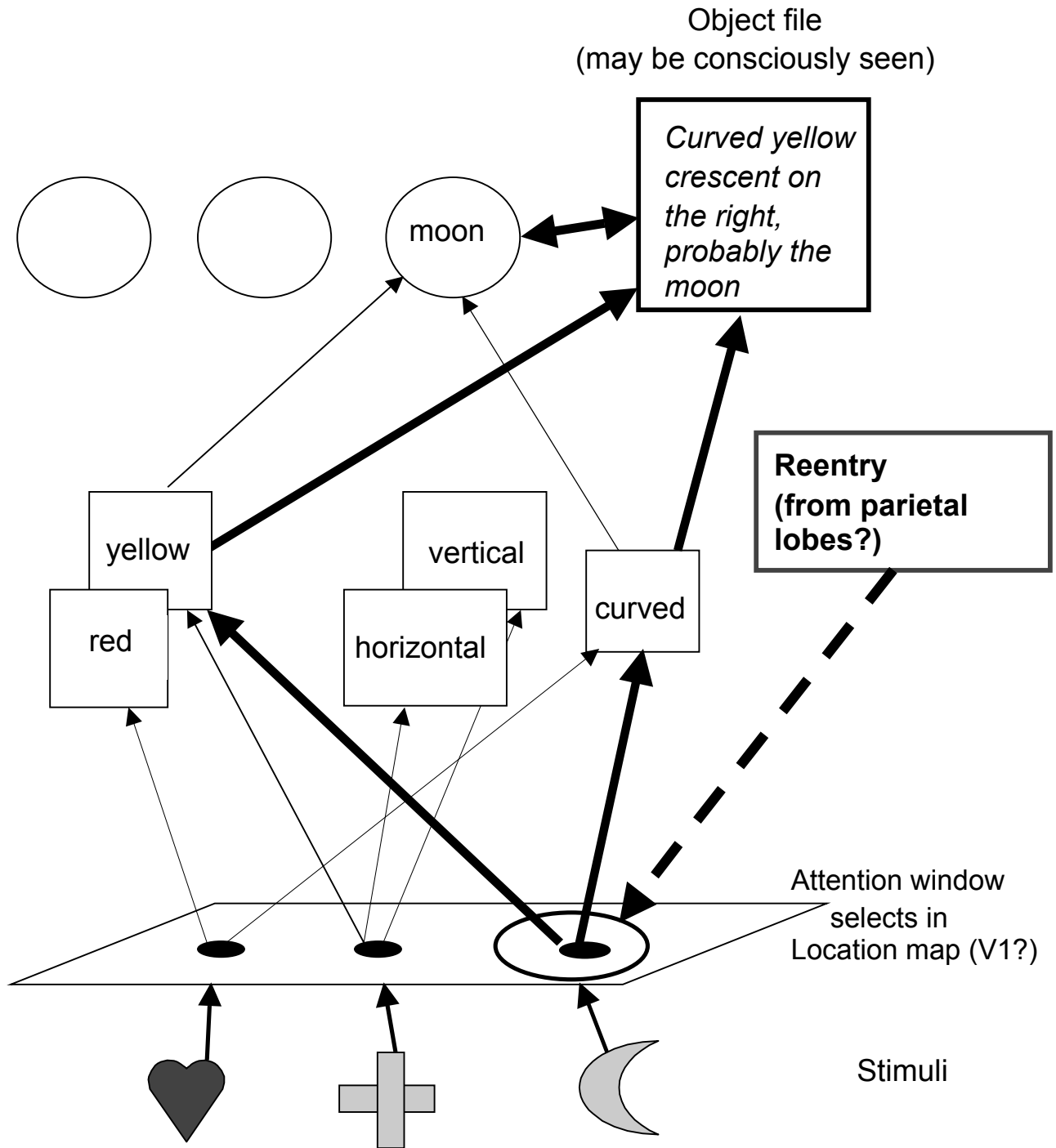
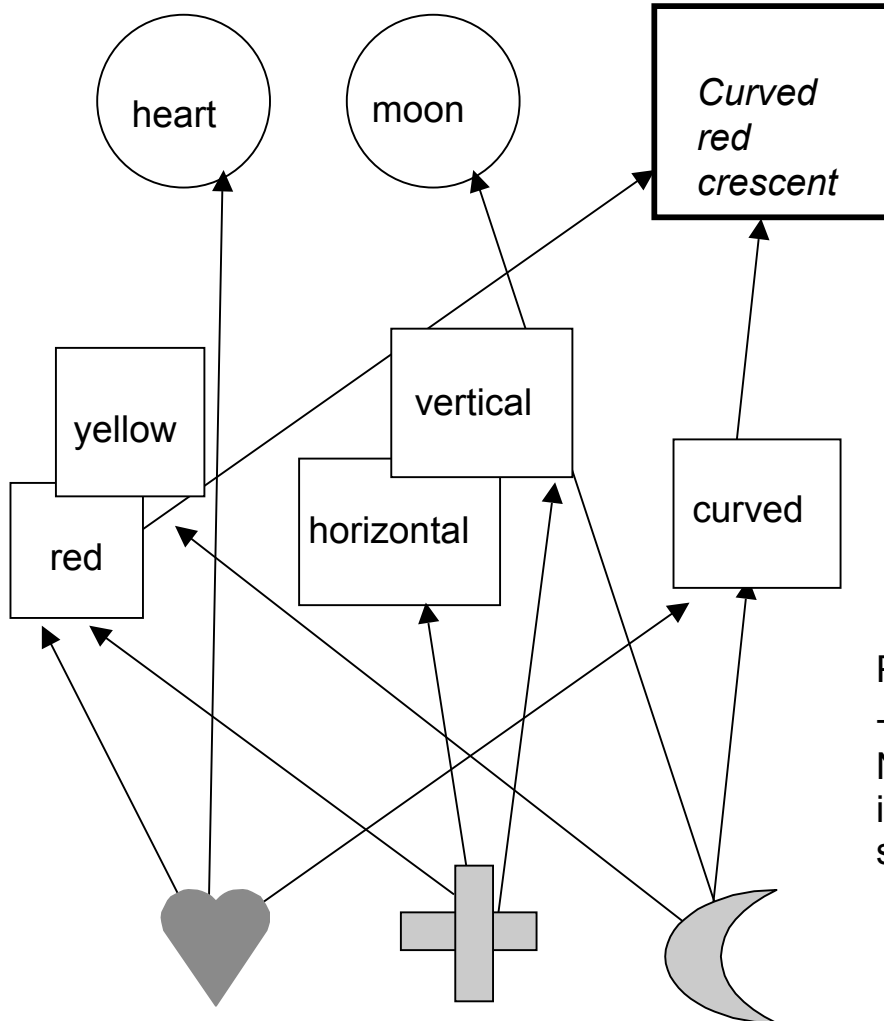


Figure 3

Balint's patient

One object file available. (Random selection of features is entered and illusory conjunction may be consciously seen.)

Features still detected; recognition network primed

Parietal map destroyed, -> no selection possible. No binding or individuation of objects in separate object files.

Stimuli

Figure 4

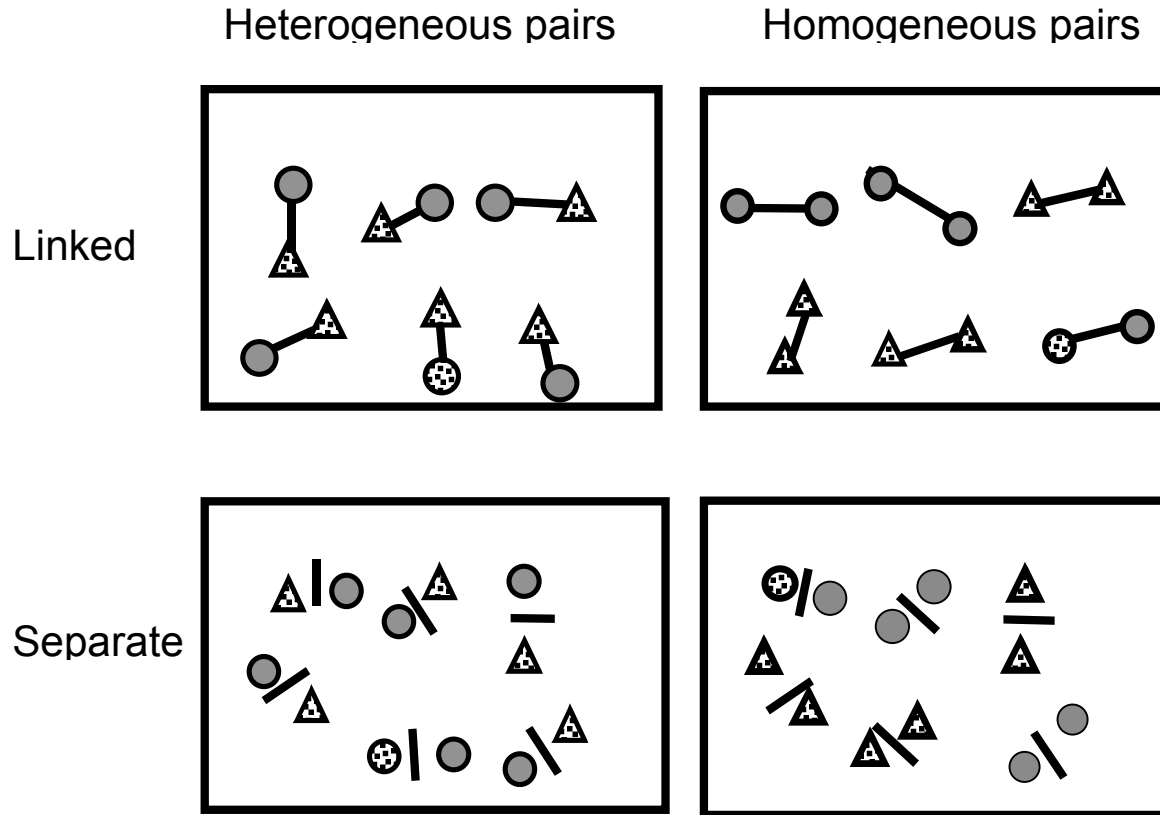


Figure 5

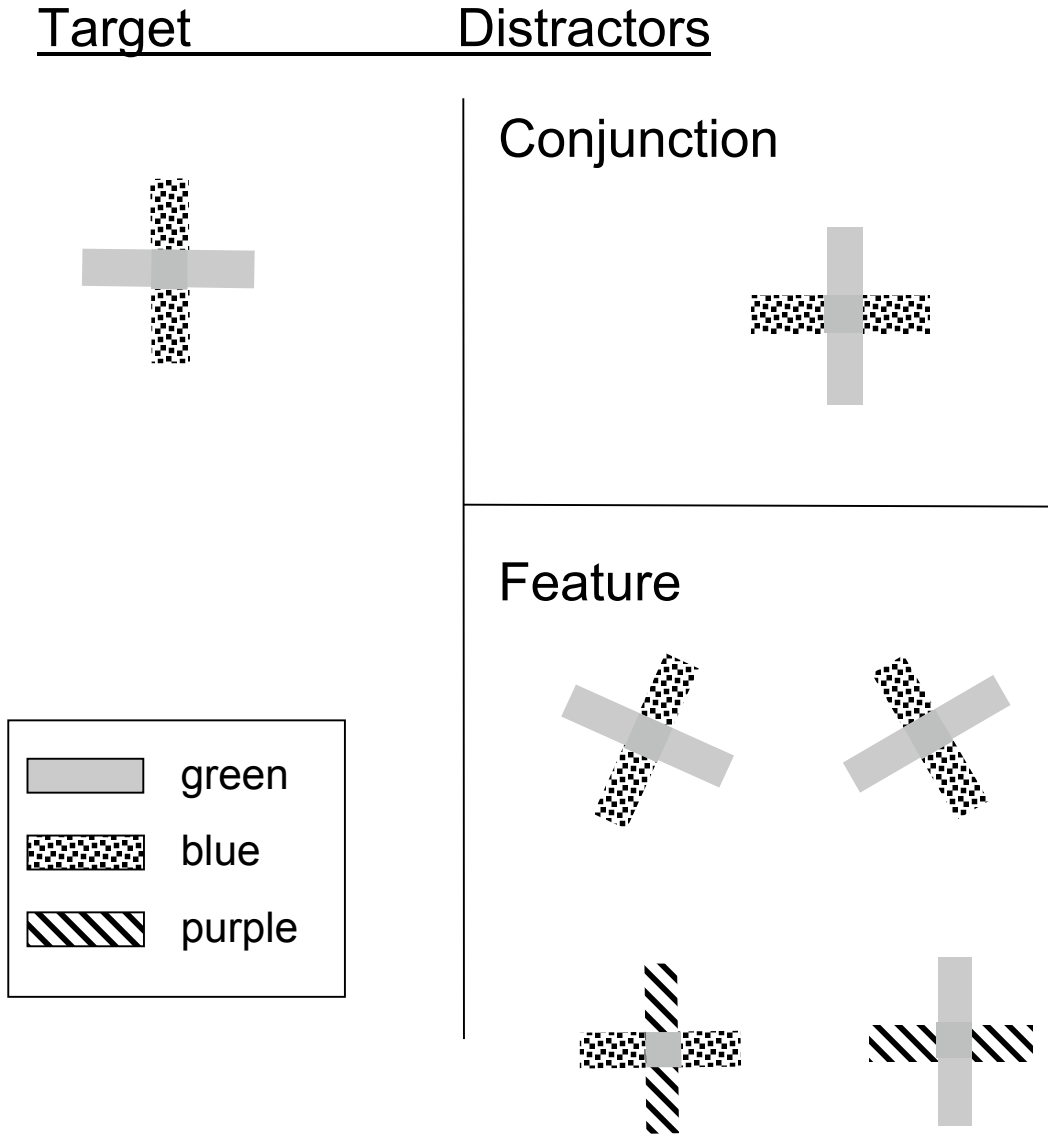


Figure 6

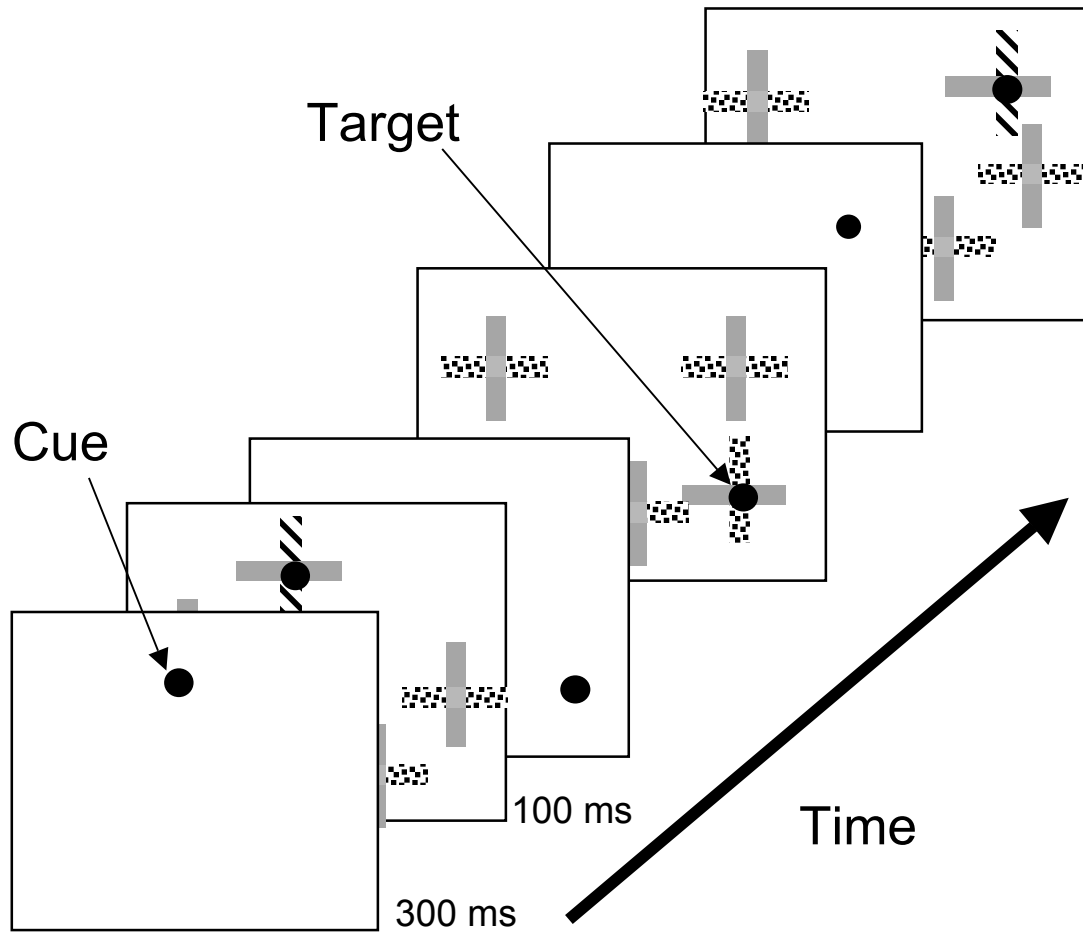


Figure 7

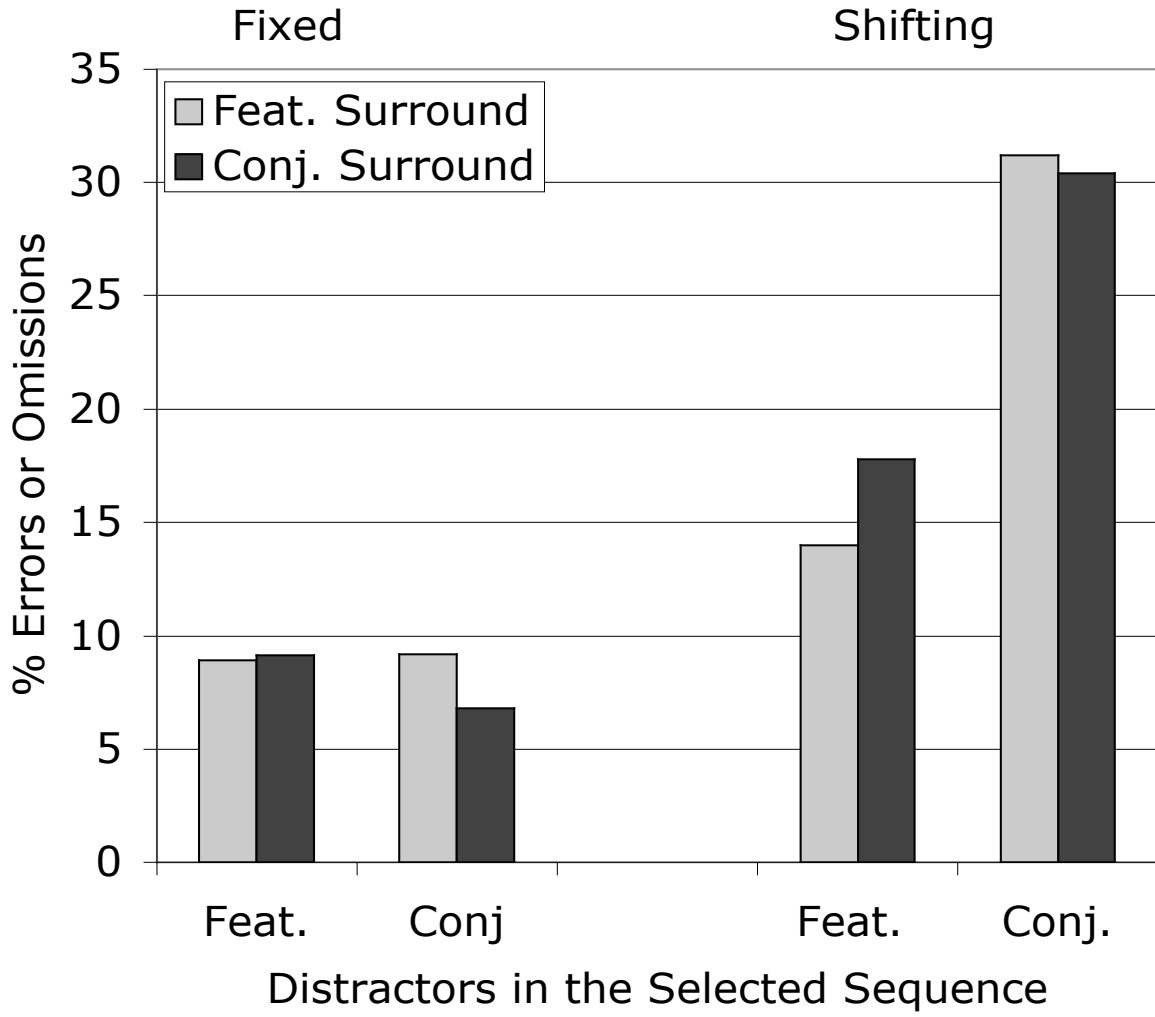


Figure 8

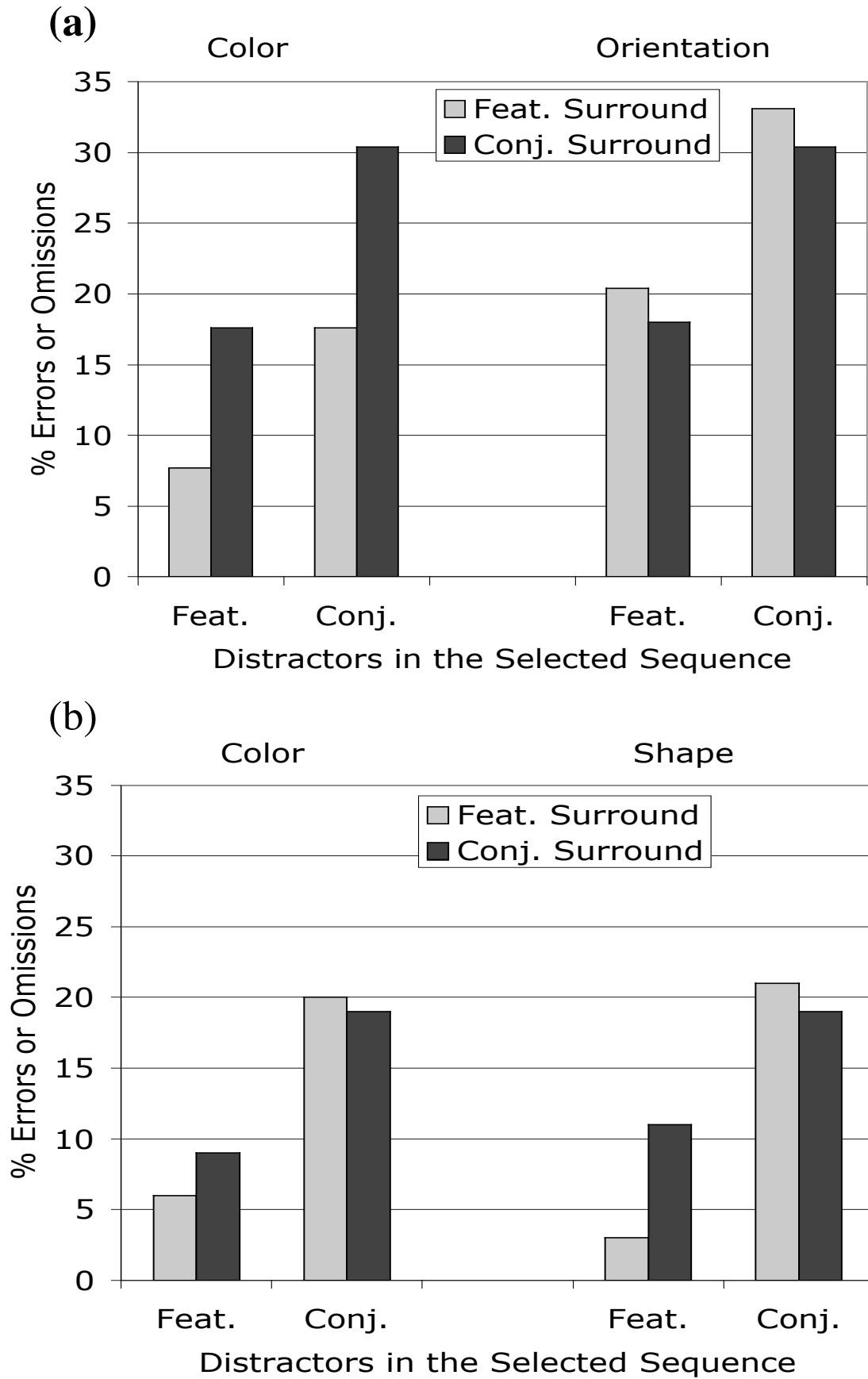


Figure 9

Global attention distributed over three objects

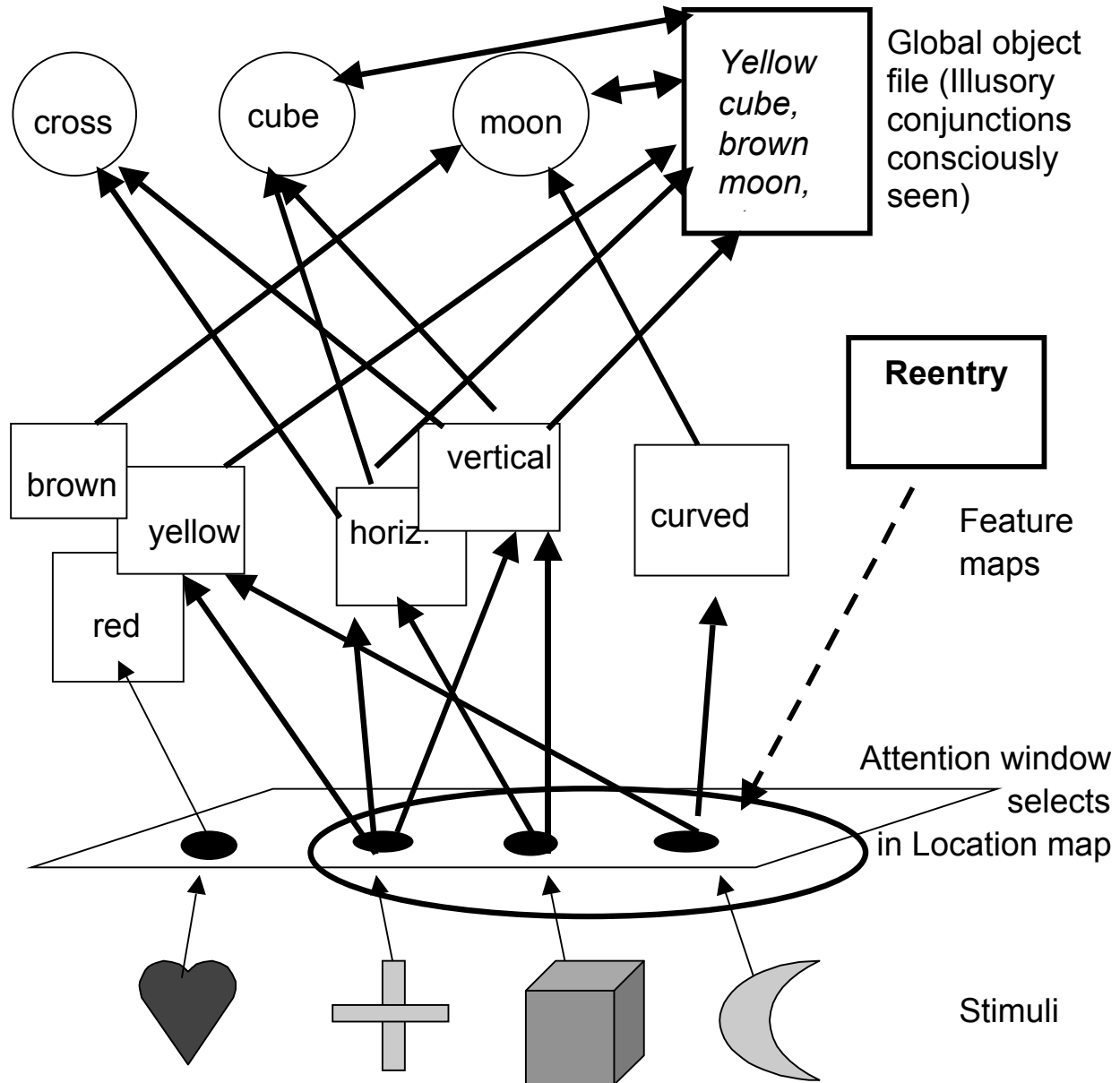


Figure 10

With divided attention: Statistical processing mode

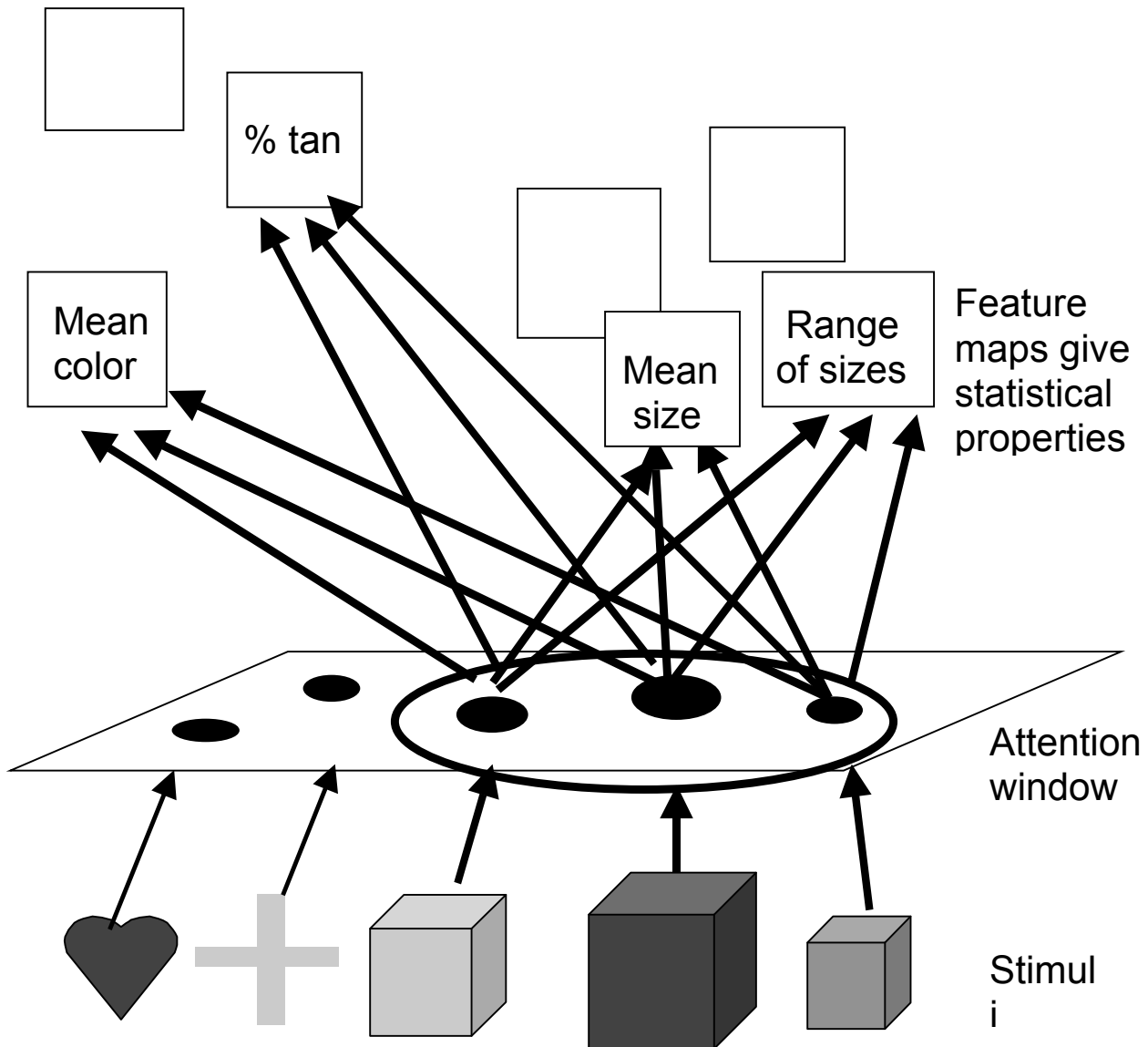


Figure 11

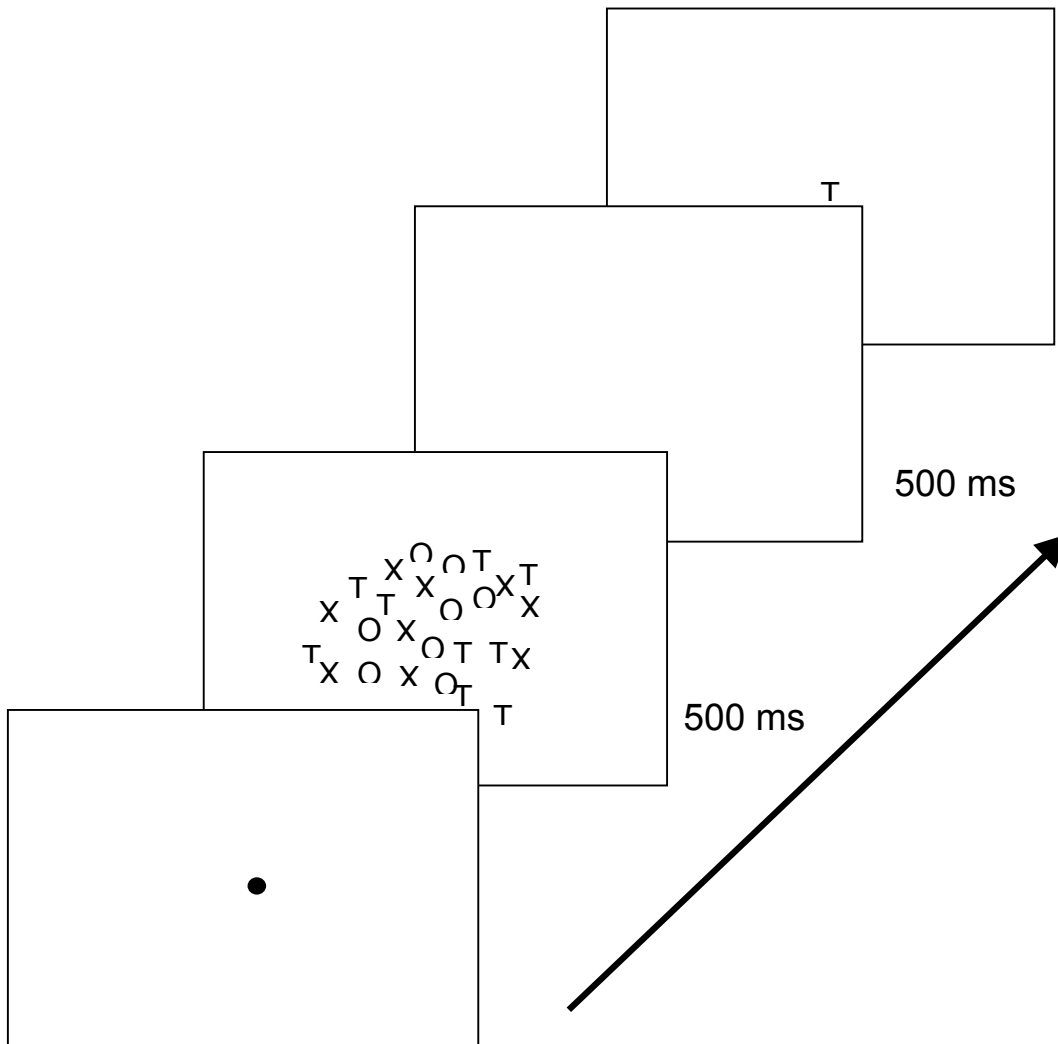
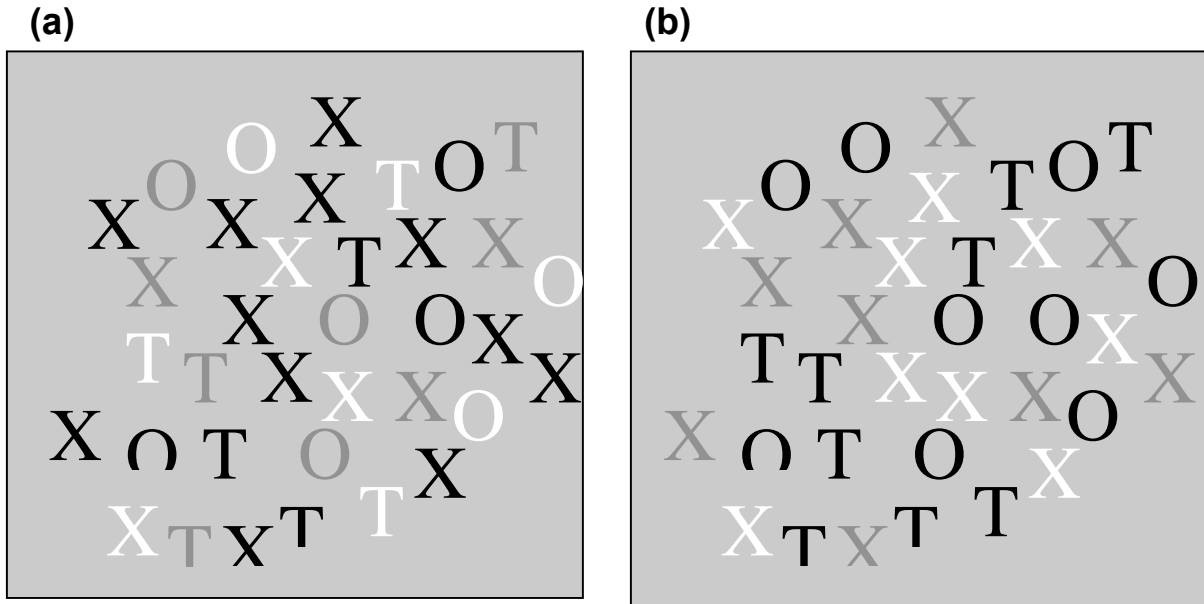


Figure 12



Key


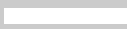
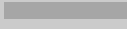
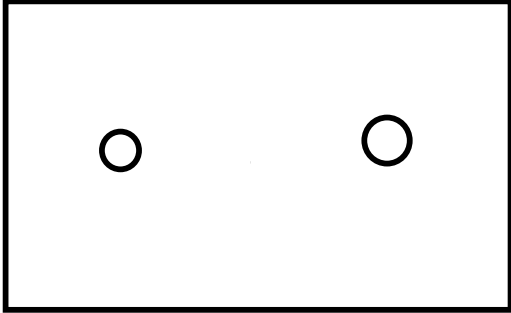
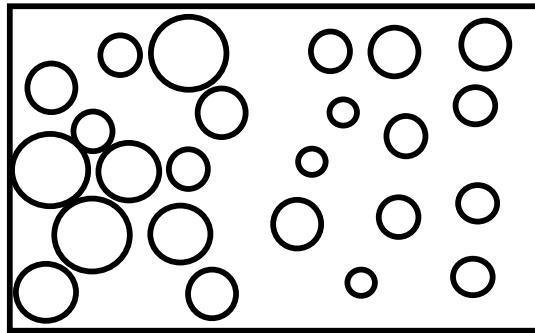
| | |
|---|-------|
|  | Blue |
|  | Green |
|  | Red |

Figure 13

Single



Mixed, for judging mean size



Homogeneous

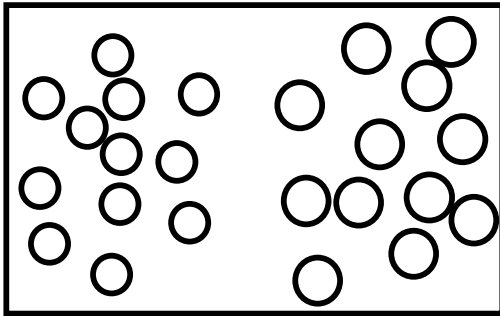


Figure 14

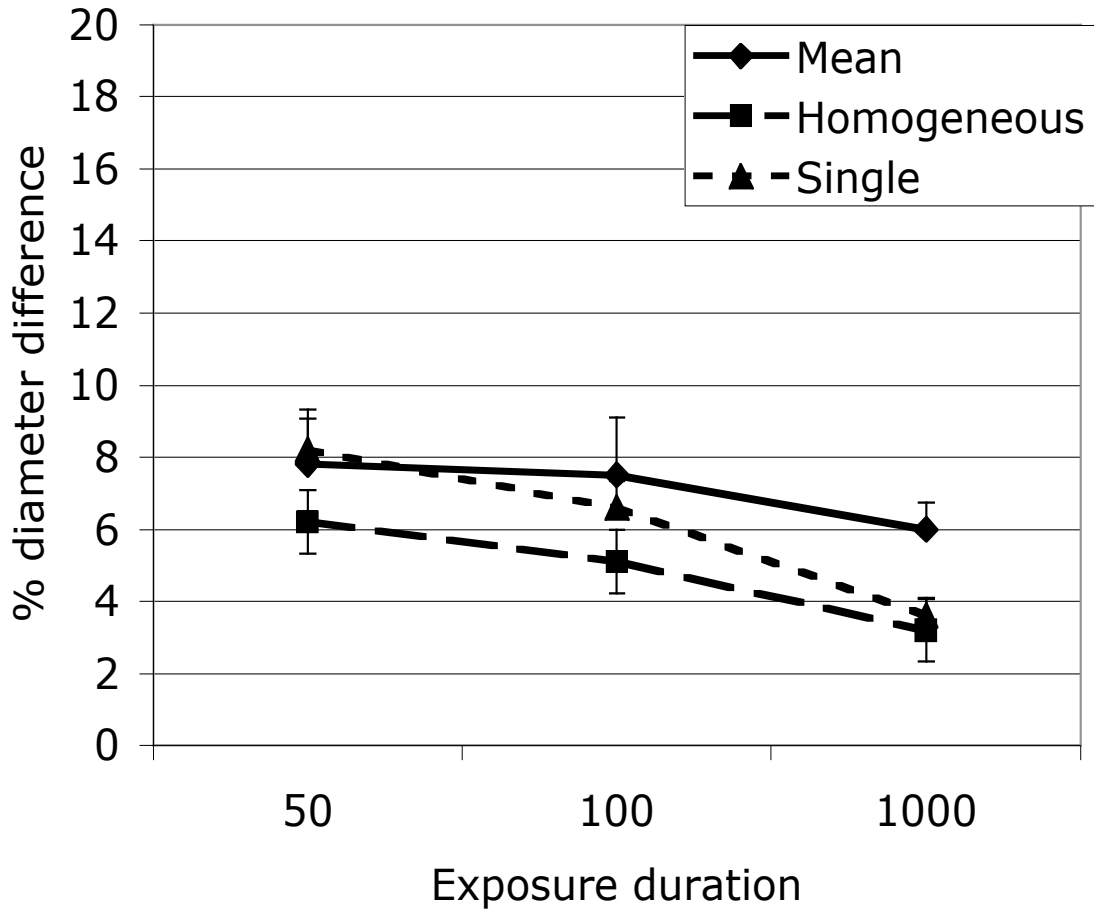


Figure 15

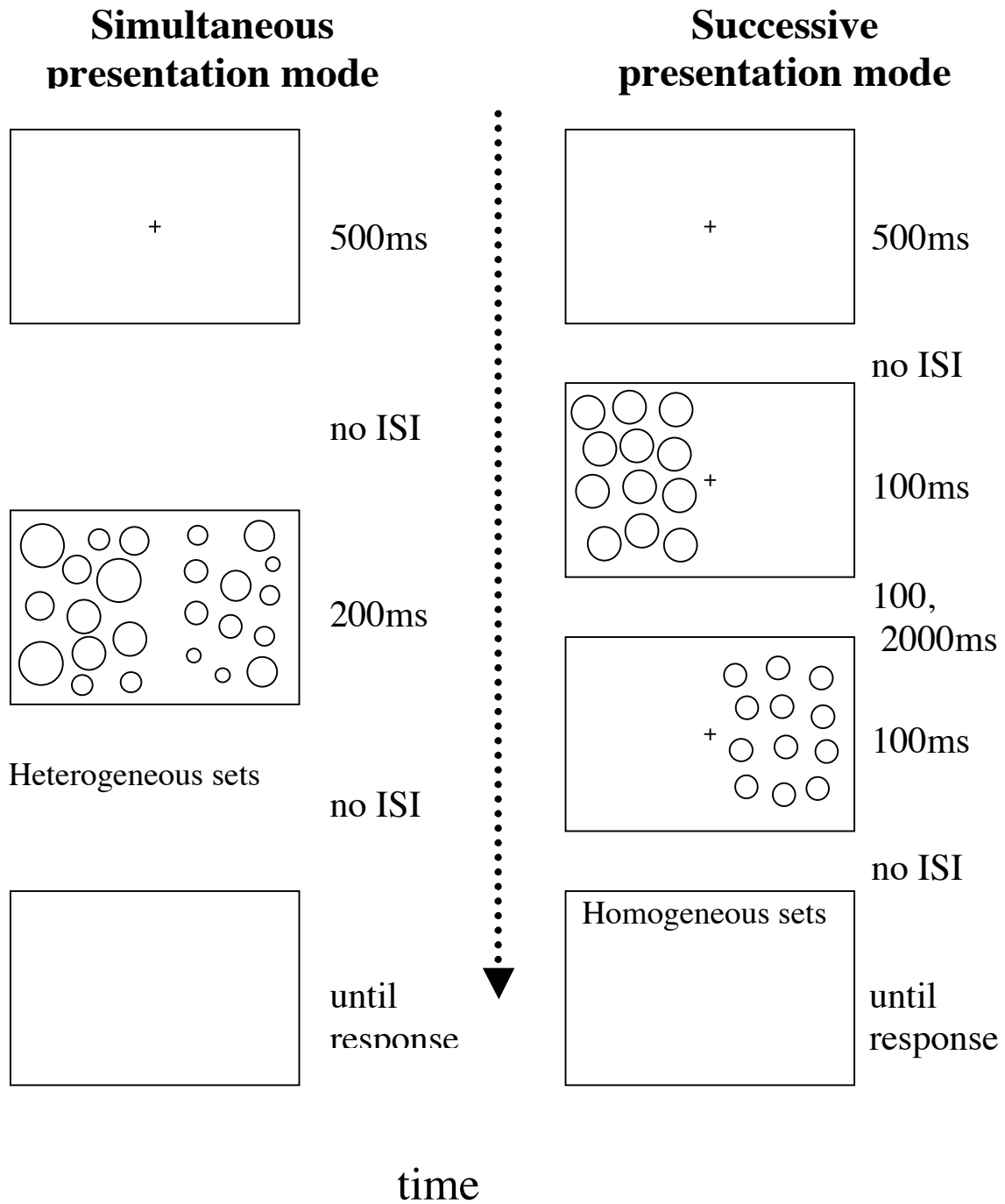


Figure 16

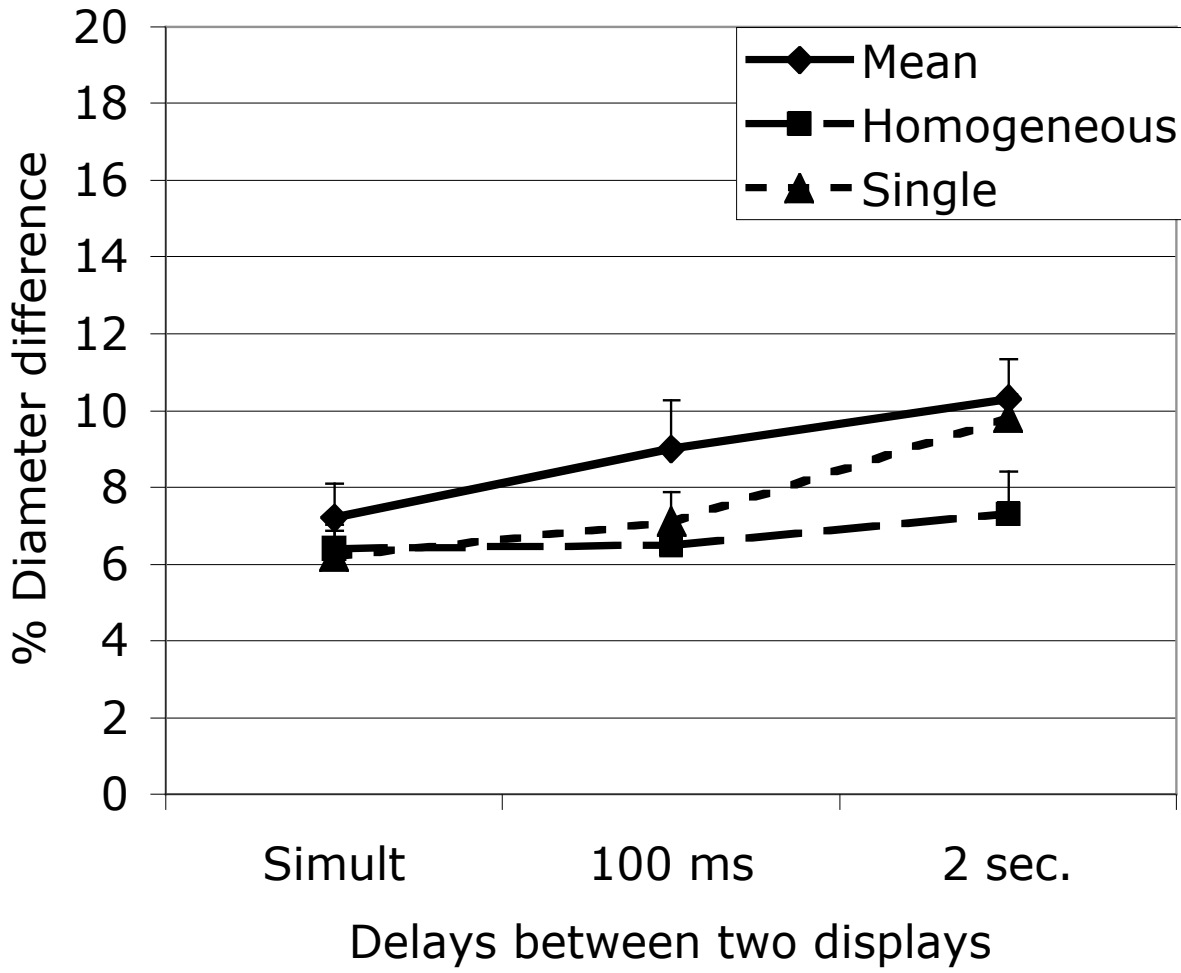


Figure 17

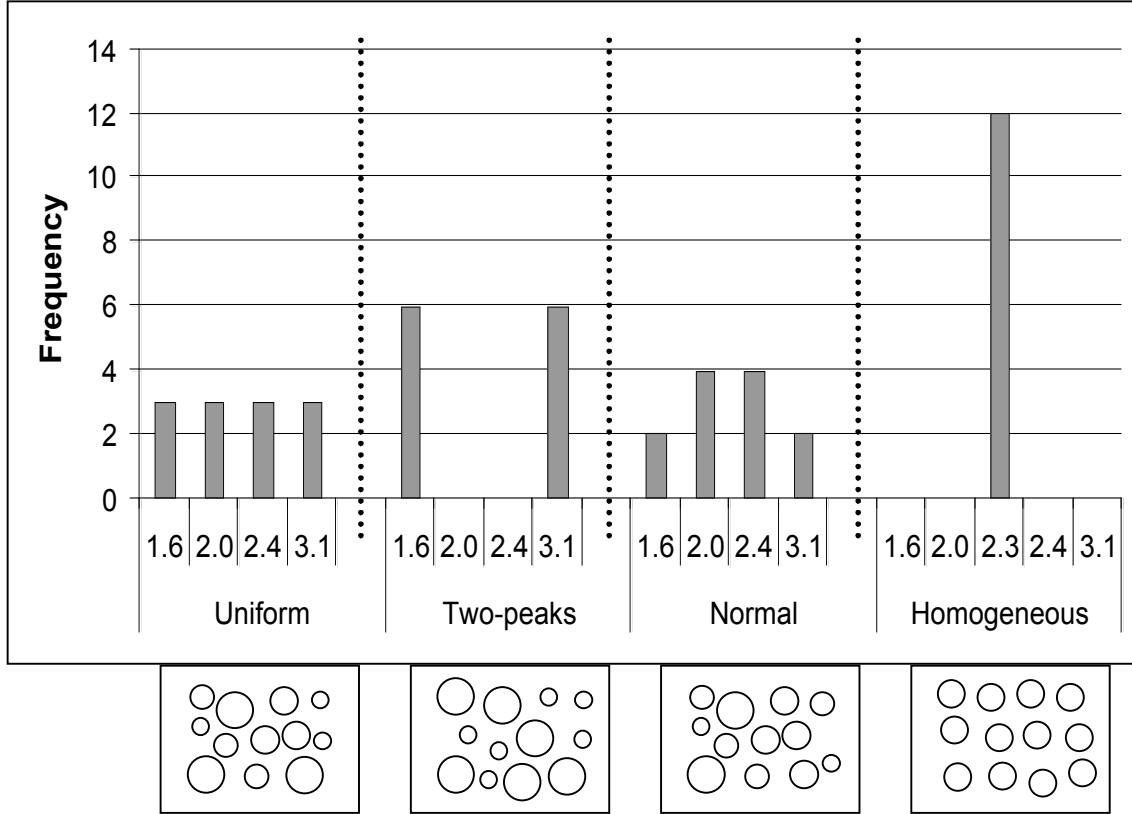


Figure 18

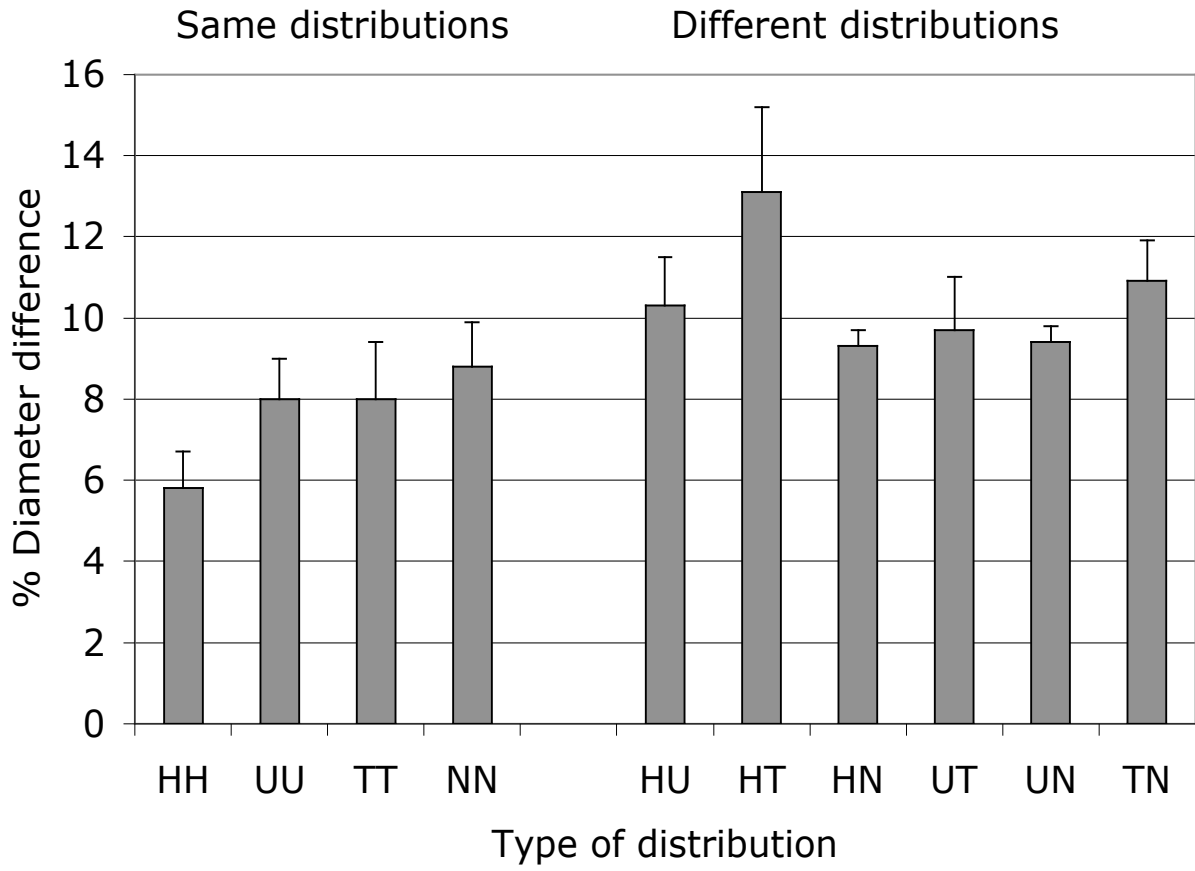
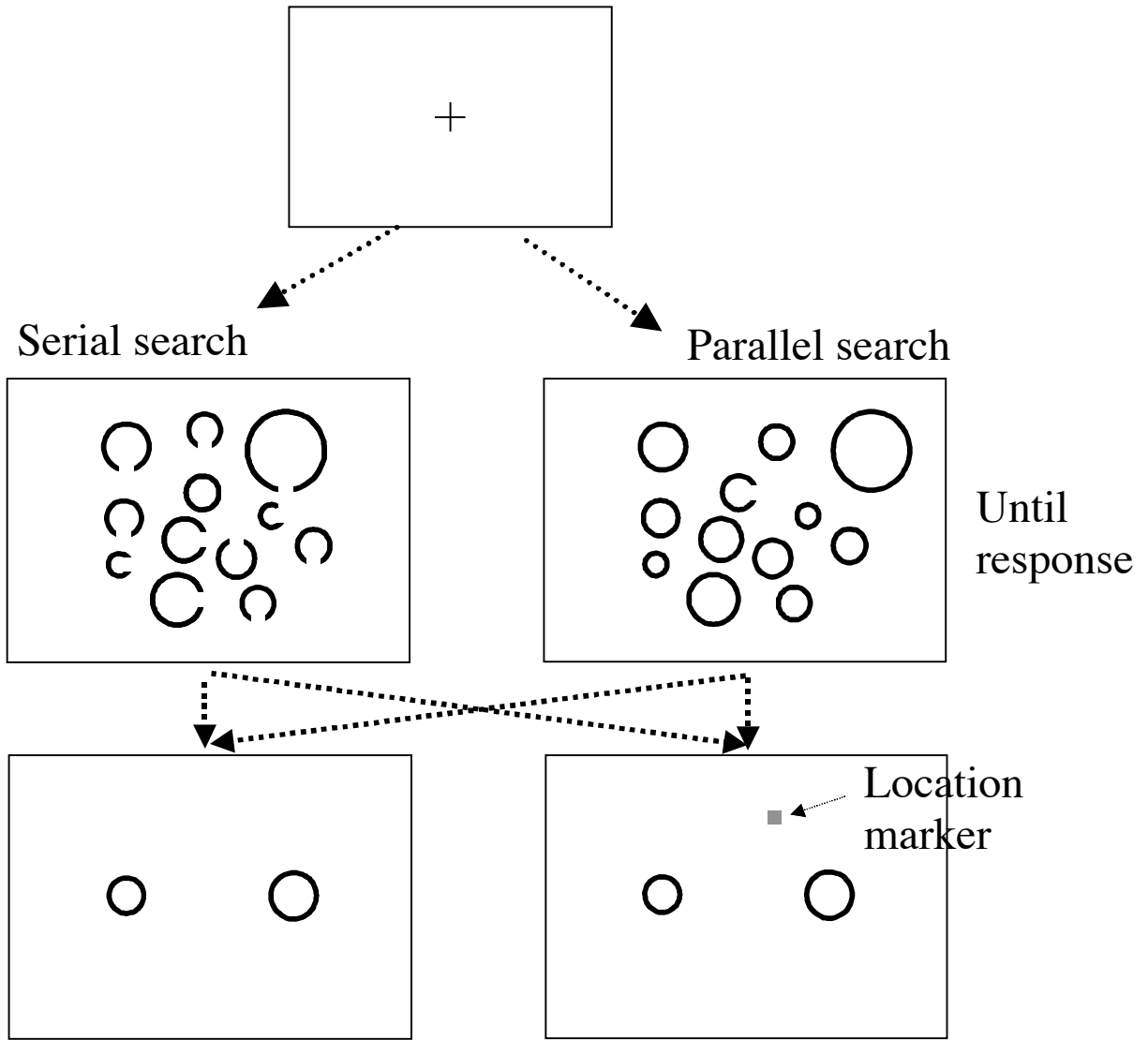


Figure 19



Mean discrimination task

Member identification task

Figure 20

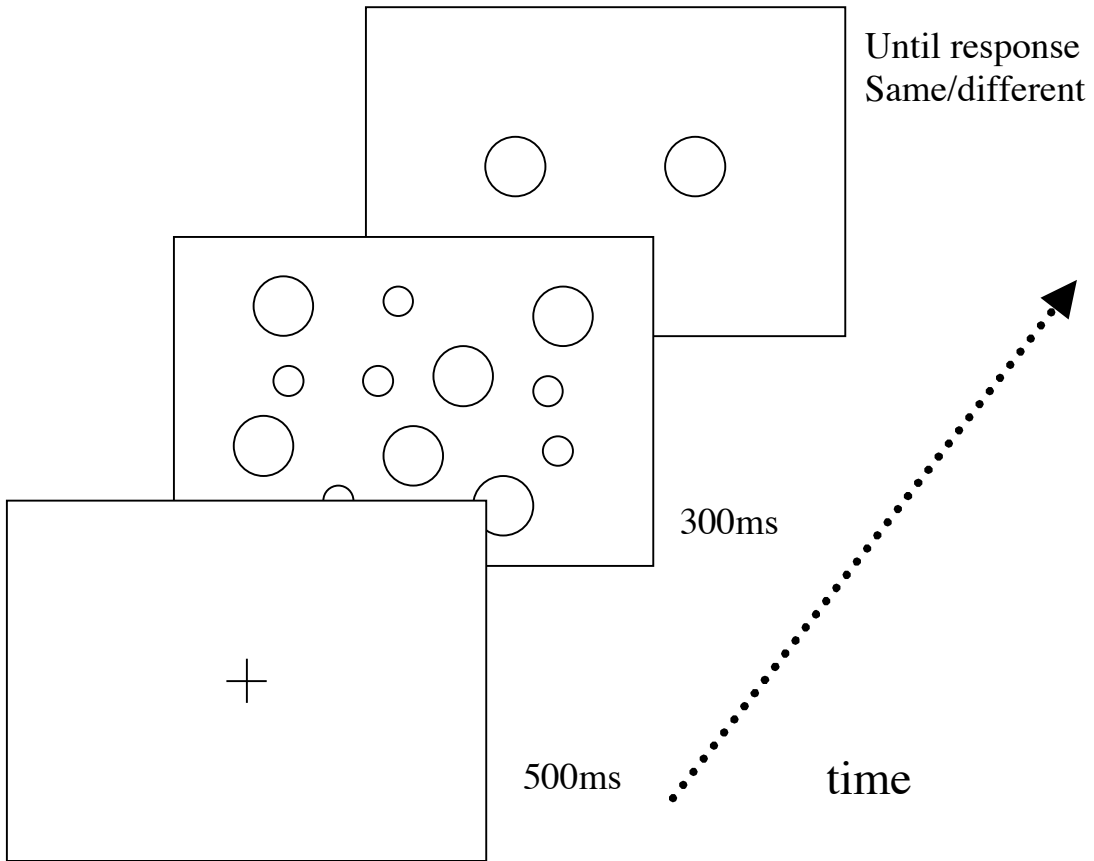


Figure 21

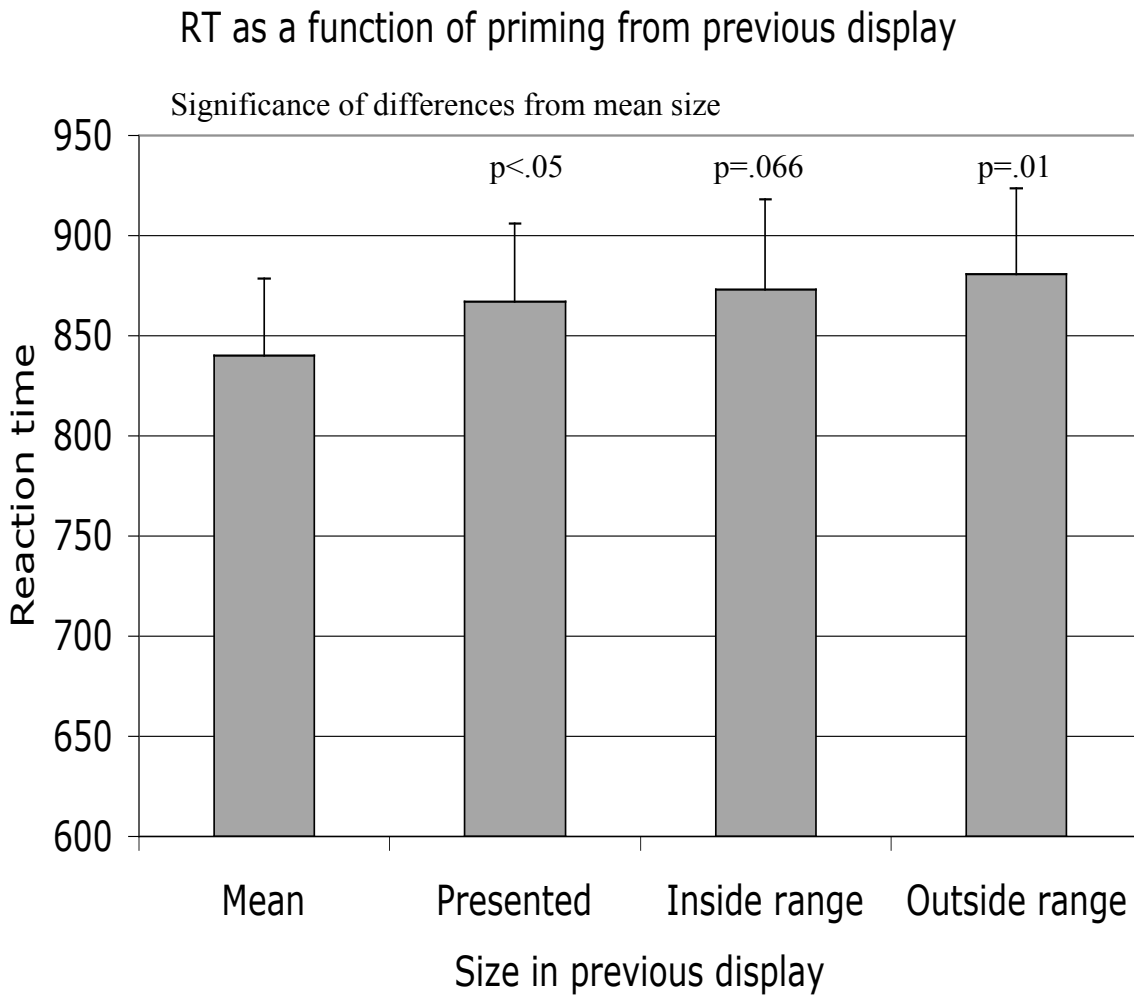


Figure 22

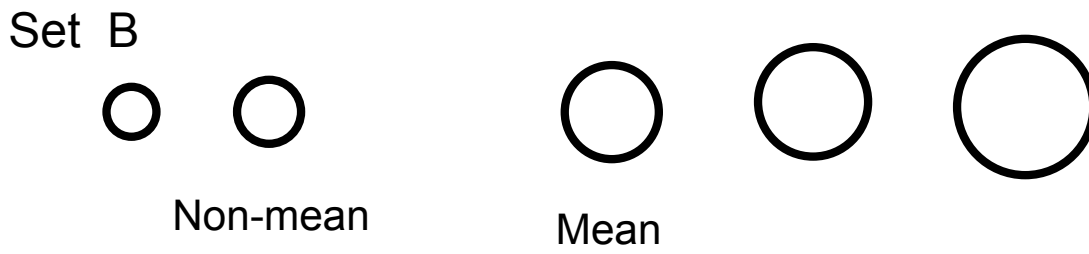
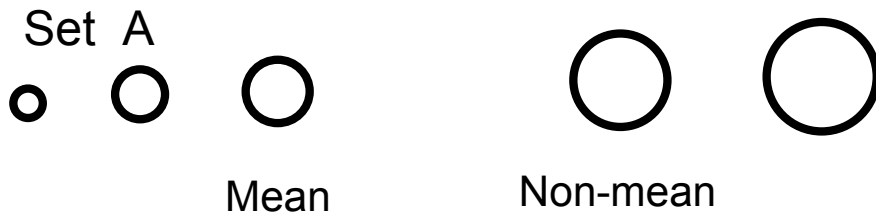
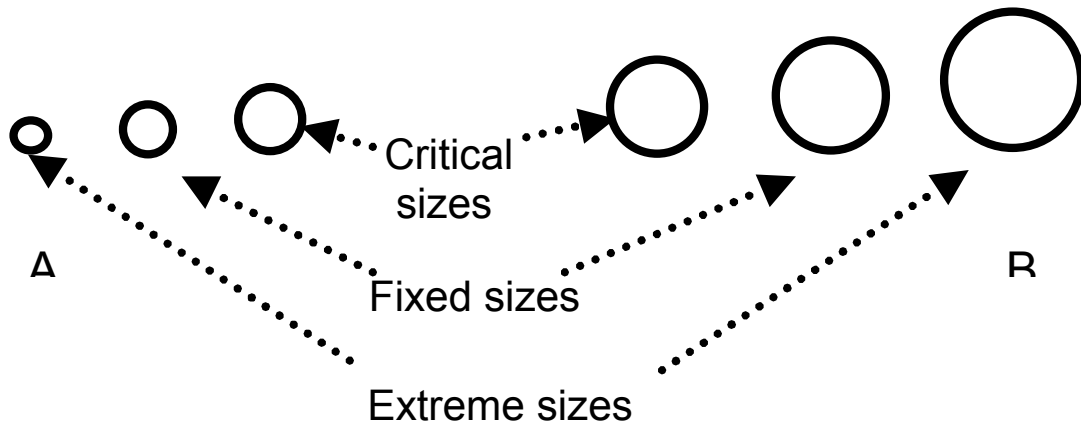


Figure 23

Without attention

