Automatic grouping of regular structures

Frouke Hermens
Laboratory of Psychophysics, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Frank Scharnowski
Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland & Laboratory of Experimental Psychology, University of Leuven, Leuven, Belgium

Michael H. Herzog
Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Introduction

To make sense of the world around us, the visual system groups elements in the stream of inputs into objects. Although there is a long tradition of research looking into the spatial grouping of static displays (for an overview, see Palmer, Brooks, & Nelson, 2003) and into how stimuli are combined across time (such as in temporal integration, e.g., Eriksen & Collins, 1967), fewer studies have looked into how information is combined across space and time (however, see Di Lollo, Enns, & Rensink, 2000; Hermens, Scharnowski, & Herzog, 2009; Herzog, Schmonsees, Boesenberg, Mertins, & Fahle, 2008; Herzog, Scharnowski, & Hermens, 2007; Otto, Ögmen, & Herzog, 2006, 2009; Oyama & Yamada, 1978; Razpurker-Apfeld & Kimchi, 2007; Turvey, 1973). One of the paradigms used to study interactions between spatial and temporal grouping is feature fusion. In feature fusion, two stimuli, which differ in one feature, are presented in rapid succession, each for a brief period of time (for example, for 20 ms each; see Figure 1A). Instead of two separate objects, participants typically report perceiving a single object whose feature is a combination of the features of the two presented objects (Efron, 1967, 1973; Kawabe, 2008; Scharnowski, Hermens, Kammer, Ögmen, & Herzog, 2007; Yund, Morgan, & Efron, 1983). For example, if two disks, a red one and a green one, are presented in rapid succession, observers report seeing only a single yellow disk (Efron, 1967, 1973; Yund et al., 1983). Similarly, if a vernier and an anti-vernier (a vernier with an offset opposite to the first vernier) are presented shortly after each other, the offsets of the verniers fuse and an almost aligned vernier is perceived (Figure 1A; Herzog et al., 2007). In feature fusion, features of objects that are presented later contribute more to the fused percept than earlier ones. Therefore, a red disk followed by a green one looks slightly greenish yellow, whereas a green disk followed by a red one has a slightly reddish tone, although the overall color is yellow. Similarly, a vernier followed by an anti-vernier results in a fused vernier with a slight offset in the direction of the anti-vernier (Figure 1A; Herzog et al., 2007).

Whether two verniers fuse depends on the context in which the verniers are presented. If the anti-vernier is not
presented in isolation, but instead, inside an array of anti-verniers (Figure 1B, left), fusion no longer takes place. This finding has been interpreted in terms of spatial grouping (Hermens et al., 2009). The anti-vernier groups spatially with the surrounding anti-verniers, which prevents temporal fusion with the preceding vernier. Temporal fusion reappears when the anti-vernier is embedded in an array of aligned verniers, presumably because the difference in vernier offsets prevents spatial grouping between the elements in the grating (Figure 1B, right; Hermens et al., 2009).

These observations exemplify the process by which the visual system binds features to objects. When a single vernier is followed by a single anti-vernier, it appears that the brain interprets the two subsequently presented stimuli to be two instances of the same object, shown at different points in time. As a consequence, it binds the features of the two verniers to this one object. For this feature binding, however, it is not a requirement that the vernier stimuli occupy the same location in space, as was shown, for example, in a sequential metaccontrast masking paradigm (Otto et al., 2006, 2009). In general, it appears that usually the visual system...
correctly binds features to objects. Instances of incorrect feature attributions, such as those found in feature fusion or sequential masking, are therefore interesting, because they reveal how this binding process might take place.

Simulations with a basic neural network demonstrated that many of the spatial and temporal grouping operations could be explained from simple neural interactions (Hermens & Ernst, 2007; Hermens et al., 2008, 2009; Herzog, Ernst, Etzold, & Eurich, 2003). Simple vernier fusion (Figure 1A) can be understood from the passive decay of information in the network and interactions with newly presented information (Hermens et al., 2009). If the same anti-vernier is embedded in a grating of anti-verniers, however, the vernier dominates the percept (Figure 1B, left). This can be understood from the interactions between neighboring neurons in the network, which suppress information inside homogeneous structures, as illustrated in Figure 1C. The resulting pattern of network activation therefore suggests that the neural network treats ensembles of similar elements as a group of which only the outer elements are highlighted. The consequence is that the anti-vernier offset is no longer evident at the center (Hermens et al., 2009), and therefore the only remaining activity at the center is related to the target vernier, which dominates offset discrimination. If, however, the anti-vernier is presented inside an array of aligned verniers (Figure 1B, right), its activity is no longer suppressed. Instead, the model seems to treat the groups of aligned verniers as separate gratings, as indicated by the highlighting of the outer elements. Because the activity of the anti-vernier is not suppressed, it fuses with the activity of the preceding vernier, just as when no surrounding elements would be present. Consequently, the model correctly predicts that discrimination performance is lower when the anti-vernier is embedded in aligned verniers than when it is surrounded by anti-verniers.

The current study investigates these grouping processes in more detail. In particular, it was tested whether fusion only occurs if the number of verniers in the target matches the number of the central group of offset elements in the mask (Experiment 1A), whether conscious access to the mask’s structure is required for fusion to occur (Experiment 1B), and whether the exact structure of the target and the mask is important (Experiment 2).

General materials and methods

Participants

Participants included the first two authors and students of the École Polytechnique Fédérale de Lausanne (EPFL). Before taking part in the experiments, observers were informed about the general purpose of the experiment and the experimental procedure, after which they signed an informed consent form. By means of an automated test (Bach, 1996), the visual acuity of each participant was determined. Participants had to reach a value of 1.0 (corresponding to 20/20) for at least one eye to take part. The students were paid 20 CHF per hour for their participation.

Apparatus

Stimuli were presented on a Tektronix 608 X–Y display equipped with a P11 blue phosphor. The display was controlled by a PC via fast 16-bit DA converters. The line elements of the stimuli were composed of dots drawn with a dot pitch of 200 μm at a dot rate of 1 MHz. The display was refreshed at 200 Hz. Participants were seated at a distance of 2 meters from the screen. A dim background light illuminated the room at approximately 0.5 lx. The luminance of the stimuli was measured with a two-dimensional dot grid using a Minolta LS-100 luminance meter.

Stimuli

Stimuli were presented at a luminance of 80 cd/m² on a dark background which had a luminance below 1 cd/m². A sequence of one (‘1V’) or five (‘5V’) target verniers and a grating mask was presented (see the illustrations below the data plot in Figure 2). The vernier(s) in the target and the grating mask consisted of two vertical segments which were each 10’ (arc minutes) long and which were separated by a vertical gap of 1’. The vernier segments of the target were always horizontally offset by 40” (arc seconds). The offset direction (left or right) was randomly chosen on each trial. Verniers in the grating mask, which always consisted of 25 elements, were either aligned (no horizontal offset) or offset in opposite direction to the target vernier(s). The target vernier(s) and the grating mask were presented for 20 ms each.

Procedure and design

On each trial, a small fixation cross was presented, followed by the stimulus sequence. This sequence consisted of one or five target verniers, whose offset direction(s) were randomly chosen (left or right), followed by a grating mask in which certain verniers were offset in the opposite direction to that of the preceding target vernier(s) (i.e. they were ‘anti-offset’). Due to the spatial superposition of the target vernier(s) and the elements of the mask, the central element(s) of the display appeared brighter than the surrounding elements and could therefore be located easily. Participants were asked to perform an offset discrimination task on these brighter looking elements by pressing one of two push buttons to indicate the offset direction (left/right). Observers were allowed to...
look at any of the brighter elements. If observers failed to give a response within 3000 ms, the next trial was started and the missed trial was repeated at the end of the block. Experimental blocks consisted of 80 trials (Experiments 1A and 1B) or 40 trials (Experiment 2). Once all conditions were presented, a short break was given, after which all conditions were repeated in reverse order. The order of conditions was randomized across participants.

Data analysis

For each block, the percentage of trials in which the observer’s responses matched the offset direction of the target vernier(s) was computed. This measure is called ‘vernier dominance’ (e.g., Hermens et al., 2009; Scharnowski, Hermens, & Herzog, 2007). If vernier dominance is above 50%, participants more often reported the offset direction of the target vernier(s). A vernier dominance below 50%, on the other hand, indicates that participants more often reported the offset direction of the anti-vernier(s) in the mask. The vernier dominance from the two repetitions of each condition was averaged into one mean value.

Model

The structure of the model is illustrated in Figure 1C. The model uses two types of neurons, organized in two layers. Half of the neurons are inhibitory and suppress the activity of neighboring neurons. The remaining neurons are excitatory and increase the activity of surrounding neurons. After input enters the model, it is first filtered by a Mexican hat type input filter. Then, neurons in the excitatory and inhibitory layer start to impose their influence on the activity of the other neurons, which is mediated by the interaction kernels: \( W_e \) (excitatory) and \( W_i \) (inhibitory). These kernels are constructed such that neurons that are nearby affect each other more strongly than neurons further apart. Moreover, the inhibitory kernel is wider than the excitatory one, which causes inhibition to spread further and faster than excitation. Furthermore, excitation and inhibition spread across the network with different time constants. Simulations suggest that the combination of the parameters for excitation and inhibition results in the highlighting of the outer elements of regular structures in the input (Hermens et al., 2008). For example, if a uniform grating is presented to the network, only the outline of the grating is highlighted (see Figure 1C, right). On the other hand, if a grating is presented in which two elements differ in their properties, such as in their luminance or size, activation in the layers of the network at the position of these different elements is increased with respect to the surround.

The parameters that are related to the input filtering and the excitation and inhibition of activation across neurons define the structural properties of the model. When these parameters are changed, the behavior of the model...
changes. For example, neurons that originally interacted no longer interact, or new interactions appear that did not exist for other settings of the parameters. Even though such changes allow a better fit to the current data, we kept these parameters at the values from the original publication (for details, see Hermens et al., 2008) to avoid overfitting of the model to the current data set. Moreover, by not changing the model’s structural parameters, we could ensure that the model continues to explain previously obtained data (Hermens et al., 2008, 2009).

Network activation was converted to predicted vernier dominance by a linking hypothesis (Hermens et al., 2009). The activity in the excitatory layer of the network is read out 20 ms after the offset of the last stimulus in the sequence and compared with a template. Activity is first matched against a template of the left-offset vernier target and then to a template of the right-offset target. The amount of overlap with each of the templates is compared (the difference is taken) and converted to a percentage value using a sigmoid function. This sigmoid function has two parameters: The horizontal position of the curve (‘shift’) and its steepness (‘slope’). The asymptotes, representing lapse rates, are fixed at 5%. The values of the two parameters of the sigmoid function were determined by minimizing the sum of the squared differences between the model predictions and the data for a small portion of the data (Experiment 1 in Hermens et al., 2009).

Experiment 1A: Only objects of the same size are temporally fused

In Experiment 1A, we show that the visual system first groups elements spatially and then integrates grouped structures temporally.

Methods

Nine observers (two females) took part in Experiment 1A, including two of the authors (FH and FS). Participants were presented with sequences of one (‘1V’) or five (‘5V’) target verniers, followed by a grating mask consisting of 25 elements (see illustrations in Figure 2). Of this grating mask, either one (‘1AV24N’, where ‘AV’ stands for anti-vernier, and ‘N’ for non-offset), five (‘5AV20N’), or all 25 elements (‘25AV’) were offset in the opposite direction from the target vernier(s). Each sequence of stimuli was presented 80 times per block. Within each block, only one combination of target and mask was tested. The offset direction (left/right) of the target vernier(s), and thereby the offset direction of the anti-vernier(s), was randomly selected on each trial. Once all conditions had been tested, their order was reversed and they were presented for a second time, resulting in a total of 160 trials per condition.

Results and discussion

Figure 2 shows the average vernier dominance, which is defined as the percentage of trials in which participants reported the offset of the target vernier(s).

For a single target vernier, performance is significantly lower for a mask containing only one central anti-vernier than for a mask with more anti-offset elements (difference contrast comparing 1V-1AV24N with 1V-25AV and 1V-5AV20N: F(1, 8) = 17.12, p = 0.0030). This high vernier dominance for the 25AV and the 5AV20N masks, which is similar to that found for a grating consisting of aligned verniers only (Hermens et al., 2009), is an indication that the anti-verniers in the masks are spatially grouped, thereby preventing temporal fusion with the preceding vernier. In contrast, the lower vernier dominance for the 1AV24N mask indicates that the single anti-vernier in the mask is temporally fused with the target.

When five target verniers precede the grating mask, performance is lowest when the grating contains five anti-verniers (difference contrast comparing 5V-5AV20N against 5V-25AV and 5V-1AV24N: F(1, 8) = 35.82, p < 0.001). This suggests that the five anti-verniers in the 5AV20N mask are grouped and because the grouped structure has the same number of elements as the preceding five verniers target, temporal fusion occurs. No such fusion appears to take place for a single anti-vernier in the mask (1AV24N) and only slightly when all 25 elements in the mask have an offset opposite to that of the target (25AV). The results are therefore in agreement with our hypothesis that the visual system first groups elements spatially and then only fuses groups that have the same number of elements.

Model simulations

We found that vernier dominance was lower when the number of offset elements in the mask matched the number of elements in the target. This suggests that the visual system first groups the elements in the mask spatially, and subsequently, fuses only groups of equal numbers of elements. Simulations with the neural network model (Hermens et al., 2008, 2009) explain why this might be the case. Figure 3A shows the activity in the excitatory layer of the model for one particular sequence, namely, the sequence of five target verniers followed by five anti-verniers in a grating mask of otherwise aligned verniers. The activation plots show that shortly after the five verniers have been presented to the network, activity is biased in the direction of the target.
This bias starts to reverse shortly after the grating mask is presented.

Similar activation plots for the other target and mask combinations are shown in Figure 3B in which the activity after 50 ms of simulated time is shown. These plots show that the model highlights places in the mask where elements differ in their spatial offset. At these highlighted areas, the offset information is maintained. If a vernier target is presented in such an area, the offsets of the target and the mask will fuse.

Figure 3C presents the predicted vernier dominance for each of the conditions, showing a pattern closely matching the experimental data (correlation, $r = 0.93, p = 0.007$; no fitted parameters). The vernier dominance tends to be slightly overestimated, probably because the model parameters from a previous study were used. In this

Figure 3. (A) Activity in the excitatory layer of the neural network model (Figure 1B) in time steps of 10 ms after the presentation of a sequence of five target verniers (for 20 ms) and a grating mask (for 20 ms; stimuli illustrated on the left) containing five anti-verniers, visible as the four dots with increased activity. (B) Activity in the excitatory layer after 50 ms of simulated time for each of the combinations of target and grating masks. In the 1V target plots, the darker two dots in the center correspond to the offset of the target vernier or the central element of the mask. In the 5V target plots, the darker four dots show activity related to the edges of the target grating or the corresponding elements in the mask. For the purpose of illustration, only the central parts of the activation map, and the illustrated gratings are shown. (C) Vernier dominance on the basis of the activity in the excitatory layer (mapped to vernier dominance using the same linking hypothesis and parameters as in Hermens et al., 2009).
previous publication (Hermens et al., 2009), slightly higher vernier dominances were obtained compared to the current study (for example, for the 1V-25AV sequence a vernier dominance around 90% was obtained previously, whereas here vernier dominance is close to 80%). This will be investigated in more detail in the discussion of Experiment 2, where the parameter values of the linking hypothesis were newly estimated using the data of both experiments.

The model accurately predicts the somewhat surprising finding that for the 5V target, vernier dominance is lower for the 25AV mask than for the 1AV24N mask. This lower vernier dominance for the 25AV mask was not found for the 1V target (i.e. the 25AV and the 5AV20N masks yielded a similar vernier dominance). Most likely, the difference for the 25AV mask for the two targets is due to lateral interactions, which causes the outer elements of the 25AV grating to interfere more with the outer elements of the five vernier target than with those of the one vernier target, simply because the distance between the interacting elements is smaller.

**Experiment 1B: Perceptual grouping and awareness**

In Experiment 1B, which was performed immediately after Experiment 1A, we tested whether the observer’s awareness of the structure of the mask was critical in obtaining the observed grouping effects. This was done by asking participants to discriminate between two of the masks used in Experiment 1A: The mask with one anti-vernier (1AV24N) and the one with five anti-offset elements in the center (5AV20N).

**Methods**

The same nine participants of Experiment 1A participated. On each trial, they were presented with either the 1AV24N mask or the 5AV20N mask of Experiment 1A for 20 ms (i.e., without the preceding vernier). Participants were asked to discriminate between the two masks, by pressing the left button if the mask with one offset element was perceived, and the right button if the mask with five offset elements was perceived. Observers completed one block of 80 trials.

**Results**

To investigate whether the awareness of the mask’s inner structure leads to stronger feature fusion effects, we plotted the performance in Experiment 1B (horizontal axis) against the difference in vernier dominance as determined in Experiment 1A (vertical axis). If awareness of the mask’s structure is required for spatial grouping of the target-mask sequence of Experiment 1, then participants with higher scores on the mask discrimination task of Experiment 1B are expected to also show larger differences in vernier dominance in Experiment 1A. For the mask with a single offset element (1AV24N), a negative correlation is expected, whereas for the mask with five offset elements (5AV20N), a positive correlation is predicted due to how the values of Experiment 1A are compared.

Indeed, for the 1AV24N mask, a negative, but non-significant correlation ($r = -0.11; p = 0.78$) is obtained. For the 5AV20N mask, a positive, but again, non-significant correlation is found ($r = 0.51, p = 0.16$). More importantly, the difference in vernier dominance is negative for all but one participant for the 1AV24N mask and positive for all participants for the 5AV20N mask, including those participants who were close to chance (50%) at the mask discrimination task. Taken together, these results suggest that awareness of the mask’s structure is not required for spatial grouping of the elements in the mask.

Note that we do not argue that observers were always unaware of the structure of the mask. In fact, some participants were fairly good at discriminating between the two types of masks. What we argue instead is that awareness of the mask’s structure is not a requirement for interactions between spatial grouping and temporal fusion to occur, which suggests that grouping occurs automatically, although some participants appear to be aware of the performed operations.

Although the model was reasonably accurate at predicting vernier dominance in the target offset discrimination task (Experiment 1A), it has problems to explain why not all participants could consistently report the number of offset elements in the mask when presented in isolation (Experiment 1B). Figure 4B shows the activation in the excitatory layer of the neural network model after presentation of each of the masks for 20 ms, followed by a blank screen for another 30 ms. As before, the network detects and highlights the inhomogeneities of the mask (places where elements differ in their offset). Because of this highlighting, it should be possible to identify the number of offset elements in the mask (one or five). It is therefore not clear at this point why all participants showed a distinct difference in vernier dominance but could not always discriminate between the types of mask used.

In Experiment 1A, two stimuli, each lasting 20 ms, were presented in succession. In contrast, Experiment 1B presented only the mask for 20 ms. The finding that some participants could not distinguish between the two types of masks in Experiment 1B, however, is not likely to be the result of this shorter presentation duration. For single
verniers, it has been found that vernier offset discrimination performance is deteriorated when the vernier is presented for 20 ms rather than for 40 ms (Morgan, Watt, & McKee, 1983), indicative of a poorer spatial resolution for stimuli presented for 20 ms or less. However, the offset sizes for which this holds are much smaller than those used in the present study. In fact, for the present offset sizes, percentages correct above 95% were previously found for single verniers presented for 20 ms or less (Hermens et al., 2009). Although the task in this study (vernier offset direction discrimination) differed from the task in Experiment 1B (discrimination of the number of offset verniers), the above 95% performance indicates that vernier offset detection is not impaired at the shorter presentation duration of 20 ms. Instead, we propose that the impairment in reporting the inner structure of the mask is caused by the flanking elements rather than the presentation duration, which is in agreement with earlier observations showing that flanking verniers strongly impair vernier offset discrimination (e.g., Malania, Herzog, & Westheimer, 2007).

Experiment 2: The role of the inner structure of the target

Experiment 1A showed that five target verniers fuse with exactly five anti-verniers in the mask, resulting in a vernier dominance for this condition well below that for masks with different numbers of anti-verniers. Experiment 2
investigates how precise the requirements on the overlap between elements in the target and mask aspects are. This experiment was inspired by the model simulations of the data of Experiment 1A, suggesting that offsets at positions in the grating where there is an inhomogeneity (neighboring elements with different offset sizes) contribute more strongly to the fused percept than elements inside a regular grating. This could mean that only the outer elements of groups of five verniers or anti-verniers are important for perception.

Methods

Six participants (four female) took part in Experiment 2, including author FH. The remaining participants were naive with respect to the purpose of the experiment. Combinations of one of three targets and one of four masks were used, resulting in a total of twelve stimulus conditions (see the illustrations in Figure 5). The target either consisted of three central verniers flanked by two aligned verniers (‘NVVVN’), two verniers flanking three aligned elements (‘VNNNV’), or five verniers (‘5V’). The masks contained similar structures of anti-verniers and aligned elements. Either two anti-verniers were flanking three aligned verniers in the center (‘AV@P3’, indicating that the anti-verniers were at position 3 from the center), five anti-verniers flanked by aligned verniers (‘5AV20N’), three central anti-verniers flanked by aligned verniers (‘3AV22N’), or 25 anti-verniers (‘25AV’). Each combination of target and mask was presented 80 times in two blocks of 40 trials each. The task of the participants was the same as in Experiment 1A, namely, to report the offset direction of the elements in the center that appeared brighter. Observers were free to base their decisions on any of these elements.

Results and discussion

Figure 5 plots the vernier dominance for the different combinations of target and mask in Experiment 2. The results demonstrate the importance of the outer elements and show that vernier elements inside the gratings (target) and inside groups of mask elements contribute much less to the fused percept. This is reflected by the fact that the targets with two outer elements with an offset (5V and VNNNV) show an almost identical pattern of results. Likewise, masks with similar offset directions at the outside of the group of five central elements also yield similar results (compare, for example, AV@P3 and 5AV20N). These observations are confirmed by statistical analyses. A 2 (targets) × 4 (masks) repeated analysis of variance compared the 5V and VNNNV targets (which have the same outer elements) for the four different masks. No interaction between the effects of the target and the mask on vernier dominance (F(3, 15) = 0.57, p = 0.64) and no main effect of the target (F(1, 5) = 0.33, p = 0.59) were found. In contrast, the main effect of the mask was significant, (F(3, 15) = 37.81, p < 0.001), showing that the
structure of the mask’s central five elements was important. If instead the effects of the NVVVN target (with aligned outer elements) and the other two targets (with outer elements with a spatial offset) are compared, an interaction between the target and the mask is found ($F(6, 30) = 2.46, p = 0.047$) as well as a main effect of the target ($F(2, 10) = 198.63, p < 0.001$) and the mask ($F(3, 15) = 43.13, p < 0.001$). A comparison of the masks rather than the targets shows a similar pattern of results. The two masks with anti-verniers as the outer two elements of the group of five central elements (AV@P3 and 5AV20N; left two bars for each target) show the same pattern of results (no interaction between target and mask: $F(2, 10) = 0.52, p = 0.61$; no significant main effect of the mask: $F(1, 5) = 3.14, p = 0.14$; however, a significant main effect of the target: $F(2, 10) = 181.64, p < 0.001$). The results for these masks clearly differ from those for the two other masks (3AV22N and 25AV; right two bars for each target), which have either aligned verniers as the outer two elements of the group of five central elements (3AV22N) or do not have an irregularity (25AV, for which spatial grouping prevents fusion).

Vernier dominance for the target with two aligned outer elements (NVVVN) is low for all masks. This can be understood by, again, assuming that only the outer elements are important when fusion occurs. Fusion for these outer elements seems to follow the pattern of results found for single verniers. It has been shown that fusion for single verniers can be described by interactions between two neurons: One coding for a left-offset vernier and one for a right-offset vernier (Scharnowski, Hermens, & Herzog, 2007). Because of the assumption that neural activity undergoes passive decay, elements that appear later in the sequence contribute more to the percept than elements presented earlier. Therefore, a vernier that is followed by an anti-vernier fuses into an almost aligned vernier with an offset slightly in the direction of the later presented anti-vernier. An aligned (neutral) vernier followed by an anti-vernier fuses into a vernier with a slightly larger offset in the direction of the anti-vernier because there is no vernier offset at the beginning of the sequence to cancel out the anti-vernier offset. The same combination rules seem to apply for the outer elements of the 5 elements targets and masks. For the NVVVN target in combination with the AV@P3 and 5AV20N masks, the aligned outer vernier fuses with the anti-vernier at the corresponding position in the mask, resulting in a fused vernier with an offset slightly in the direction of the anti-vernier. As a consequence, a vernier dominance of less than 50% (i.e., the offset of the anti-vernier is reported more often) is found. However, if the NVVVN target is combined with the 3AV22N mask, the aligned element in the target fuses with an aligned element in the mask, and vernier dominance is around 50%. For the combination of the NVVVN target with the 25AV mask, a vernier dominance below 50% is obtained. If all information inside the mask is suppressed, it would have been expected that vernier dominance for this condition would have been at 50% because the suppressed inside of the 25AV mask results in no remaining signal for the outer element of the NVVVN target to fuse with. Possibly, however, the edges of the 25AV grating influence the edges of the target slightly (similar to what appeared to take place in the 5V-25AV condition of Experiment 1).

The above considerations are confirmed by the plots showing the activation of the excitatory layer of the model (Figure 6A). For the NVVVN target, the anti-verniers in the mask almost always dominate, except when one of the outer two elements of the inner five elements of the mask are aligned. For the VNNVN target, the anti-verniers of the mask dominate slightly when these are present at the outer two positions of the central five elements. Otherwise, the verniers at the outer positions of the target dominate. The pattern is similar for the 5V target.

Network activation was converted to predicted vernier dominance using the linking hypothesis described earlier (Methods section; Hermens et al., 2009). Although the predicted and observed data match reasonably well ($r = 0.75, p = 0.005$, no fitted parameters), some discrepancies exist for the NVVVN target and the 3AV22N and AV@P3 masks (Figure 6B). As for Experiment 1A, vernier dominance is overestimated for these target and mask combinations. To investigate the possible reasons for this overestimation, additional simulations were performed.

Differences in baseline performance and strategies

For the simulations so far, we used the same parameter values for the model as in previous studies (Hermens et al., 2008, 2009). Particularly for Experiment 2, these parameter values did not appear to provide an optimal model fit. Here, we investigate two reasons for why the model predictions did not match the behavioral data for all conditions: (1) differences in baseline performance between participants and (2) a possible alternative strategy for the targets containing five elements.

Differences in baseline performance

Some of the stimulus conditions in the current experiments were replications of conditions used in a previous study (Hermens et al., 2009). For these conditions, however, the present study revealed a vernier dominance that differed slightly from that obtained previously, although the general pattern of results was replicated. For example, the ‘V-25AV’ combination yielded a 90% vernier dominance in the previous study, whereas a 80% vernier dominance was found here. Similarly, the ‘V-AV24N’ combination previously yielded a vernier dominance that was slightly higher than 60%, whereas in the present study vernier dominance equaled 60%. These differences suggest that baseline vernier dominance can vary slightly across participants. To investigate whether the overestimation of the vernier dominance for the
present data was due to an overall lower vernier dominance compared to the data for which the original parameters estimates were obtained, the parameters of the linking hypothesis (the mid-point and slope of the sigmoid function linking network activation and predicted vernier dominance) and the read-out time were newly fit to the current data. Parameters, such as the widths of the inhibition and excitation kernels change the behavior of the network qualitatively as well as quantitatively and were therefore left unchanged. Keeping the original values of these structural parameters also prevents the overfitting of the model to the current data set.

Figure 7B shows the predicted vernier dominance for each of the conditions for the best fitting parameters (slope of the linking function: 0.21, midpoint of the linking function: 0.40, read-out time: 61 ms), which were obtained by a simplex search of the minimum sum of squared differences between the observed and the predicted data. The model fit improved with the new set of parameters ($r = 0.80, p < 0.001$; 3 fitted parameters; whereas an overall correlation of 0.75, $p < 0.001$ was obtained without newly fitting the parameters of the linking hypothesis). For example, the overestimation of the vernier dominance for the AV@P3 mask was greatly reduced and

Figure 6. (A) Activation of the excitatory layer of the neural network model after 50 ms of simulated time for different combinations of targets (rows) and masks (columns; targets and masks illustrated next to the activation plots). (B) Predicted vernier dominance for each of the target and mask combinations.
the predictions for the NVVVN target improved. However, some of the predictions for Experiment 2 are still at odds with the experimental findings. For example, the vernier dominance for the NVVVN target–AV@P3 mask combination is overestimated, whereas the vernier dominance for the VNNNV-25AV and 5V-25AV combinations is underestimated. Because the incorrect model predictions were specific to Experiment 2, in which not all target elements had the same offset, we investigated an alternative explanation, namely, that participants focused on the outer elements of the target rather than considering all five elements.

**Strategies**

A possible reason for the less than perfect data fit for Experiment 2 could be that participants, covertly or overtly (by fixating slightly left or right of the fixation point) attended to the outer elements of the target instead of all five elements. The consequences of such a strategy for the model predictions can be investigated by assuming that participants used a template containing a single vernier at the outer position of the 5V target rather than a template of five verniers when deciding about the offset direction of the target (see illustrations below Figure 7). In a second simulation, we therefore changed the templates from a 5V and a 5AV grating to a 1V@P3 and a 1AV@P3 template (i.e., a vernier or an anti-vernier at the left or right third position from the center, at the outer position of the target). Figure 7C shows the predicted vernier dominances for the alternative templates. The data fit clearly improved now showing a correlation between the observed and predicted data of 0.89 ($p < 0.001$; 3 fitted parameters: slope of the linking function: 0.25, midpoint of the linking function: 0.31, read-out time: 46 ms). However, some discrepancies between the observed and predicted data of Experiment 2 are still present, in particular when the difference between the 3AV22N and the 25AV masks is considered. In addition, changing the templates worsened the fit of the model to the data of Experiment 1, in particular the prediction for the 5V-25AV combination. We therefore considered another possibility, namely, that strategies varied across participants. For

**Figure 7.** (A) Summary of the observed vernier dominance for the two experiments. (B) Predicted vernier dominance with the newly fitted parameters of the linking hypothesis and read-out time. (C) Predicted vernier dominance assuming all participants determined the vernier offset of the five elements target by inspecting the left or right outer element of the target. The parameter values of the linking hypothesis and the read-out time were newly fit for this hypothesis. (B and C) Below the graphs, illustrations of the used templates are provided. For purpose of illustration, the positions of the verniers in the target are shown in light gray for the 1V-1AV@P3 template (C).
this, we assumed that some participants considered all five verniers in the target, while others focused on the outer elements. In addition, it could also be that the strategy varied across experiments. In Experiment 1, participants were presented with 1V and 5V targets, whereas in Experiment 2, only 5 elements targets were used. This might have led to more participants in Experiment 2 focusing on the outer elements of the target than for Experiment 1. Additional simulations, in which the parameters were determined separately for each participant and each experiment suggested that this was indeed the case. For Experiment 1, 4 out of 9 participants showed a better date fit for the 5V and 5AV templates. In contrast, for Experiment 2 only for one out of 6 participants the data was better predicted with the 5 elements templates.

Although calibration of the linking hypothesis and the change of templates clearly improved the fit of the model to the data, some discrepancies between the observed and predicted data remained. For example, the vernier dominance for the AV@P3 and 5AV20N mask in combination with the NVVNV target is still overestimated by approximately 20%, and the vernier dominance for the 25AV mask in combination with the VNANN and 5V target is underestimated by the same amount. Generally, however, the pattern of results is well reproduced by the model. Possibly, the descriptive power of the model could be improved by including additional features to the model, for example, by assuming that the interaction strengths between neurons and the corresponding kernel widths differ for central and peripheral locations, or by assuming that instead of using a fixed offset size for the template, the overall orientation of the network activation was estimated (c.f., Hess, Barnes, Dumoulin, & Dakin, 2003). For such extensions of the model, however, it is important to demonstrate that the extensions not only improve the data fit to the current set of data but also account for the large range of existing findings in backward masking (Hermens et al., 2008) and feature fusion (Hermens et al., 2009) that the current version of the model already accounts for.

General discussion

Previous investigations of visual grouping often approached grouping across space and grouping over time separately. Recent findings have suggested that such an approach is not tenable: spatial and temporal aspects should be studied jointly (Herzog, 2007). Here, we investigated visual grouping across space and time using a feature fusion paradigm. Experiment 1A showed that vernier fusion only occurs when the size of the target (the first stimulus) matched the size of the group of elements in the center of the masking stimulus (the second stimulus). The results of Experiment 1B suggested that determining which elements to spatially group, does not require participants’ awareness. Experiment 2 demonstrated that this group of elements in the mask is defined by the outmost elements with an offset different from the surround. These results confirm our earlier conclusion that fusion across time depends on grouping across space (Hermens et al., 2009).

Interestingly, the ability of participants to report the structure of the pattern mask did not affect fusion. In fact, similar feature fusion was found for those participants who could and for those who could not report how many offset elements were in the mask, when presented in isolation. This suggests that spatial grouping operations are largely automatic. These findings are in line with observations showing that perceptual grouping does not require focused attention (Kimchi & Razpurker-Apfeld, 2004).

It is not surprising that some participants had difficulties reporting the inner structure of the mask, as such effects have been reported before. For example, offset discrimination performance is higher for the outer elements than for elements inside a grating (Sharikadze, Fahle, & Herzog, 2005). Our findings resemble findings in foveal and peripheral crowding where grouping was shown to be an important factor. In these studies, targets were flanked by distractors. Performance improved when the distractors grouped with each other (e.g., Banks & White, 1984; Livne & Sagi, 2007) and, particularly, when they ungrouped from the target (foveal: Malania et al., 2007; peripheral: Saarela, Sayim, Westheimer, & Herzog, 2009). A similar effect is found in feature inheritance, in which a vernier preceding a grating of five aligned verniers is not visible. However, the vernier’s offset is perceived at the outside of the grating (Herzog & Koch, 2001). If the central grating elements are offset, their offset does not influence performance, whereas offsets elements at the grating’s outer positions can strongly influence perception and performance (Herzog & Koch, 2001; Sharikadze et al., 2005).

Whereas a simple neural network could well explain the findings of Experiment 1, some discrepancies between the observed and predicted data were found for Experiment 2. The model’s predictions improved by assuming that participants focused on the outer elements of the five elements targets rather than by considering all five elements. When assuming that the processes in the model capture the participants’ performance, this suggests that the offset direction of the five elements targets was determined by inspecting the edge elements only. Such a conclusion is consistent with earlier observations suggesting that information at the outer positions of regular structures is more easily available than information inside a homogeneous group of elements (e.g., Herzog & Koch, 2001; Sharikadze et al., 2005). A good strategy would therefore be to focus on the outer two elements of the five elements targets, for which the offset information can be more easily accessed. The conclusion from the model that participants appear to focus on the outer elements of the
target, however, does not mean that the context in which the elements were presented did not play a role. This is illustrated by the observation that the structure of the mask strongly affected the interaction with the target. For example, for the 25 elements mask (25 AV), the offset of the mask’s elements at the position of the outer two of the five elements of the target did not influence offset discrimination. For the mask with only five anti-verniers (5AV20N), the offset of the mask element at the outer positions of the target, however, strongly influenced offset discrimination. Therefore, the spatial grouping of the mask’s elements determines how the outer elements of the target are temporally fused (see also Hermens et al., 2009).

The present data, in combination with earlier findings (Hermens et al., 2009), suggest that feature fusion can be used as a tool to study perceptual grouping. For example, it was shown that fusion over time started to dominate as soon as grouping across space weakened (Hermens et al., 2009). The feature fusion method has the advantage over previous methods to study perceptual grouping in that it allows for determining how elements in scenes are grouped without explicitly asking participants to report about this grouping. Instead, participants simply perform a vernier offset discrimination task. Moreover, vernier fusion provides a measure of the strength of spatial grouping, without having to measure the strength of one grouping factor (e.g., by proximity) by varying another (e.g., by similarity). There are methods to overcome this problem, but they rely on a model fit to the data (Kubovy, Holcombe, & Wagemans, 1998). Feature fusion, instead provides a direct method. For example, the strength of grouping as a function of proximity was obtained by varying the distance of the anti-verniers in the 25AV mask, showing that grouping by proximity breaks down at a spacing between the elements of the 25 AV grating of approximately 600” (Hermens et al., 2009). Similarly, the effects of stimulus similarity on grouping could be measured (Hermens et al., 2009).

Feature fusion provides an example of the binding problem. The binding problem refers to the task of the visual system to determine which features, such as colors, belong to which objects in the scene (Engel, König, Kreiter, Schillen, & Singer, 1992; Singer & Gray, 1995; Treisman & Gelade, 1980). This is not a trivial task because aspects, such as the color and location of an object, are processed in different parts of the brain. Situations can be found in which the visual system misattributes features to objects and illusory conjunctions are found. These illusory conjunctions can occur in the spatial domain (e.g., Treisman & Schmidt, 1982) and in the temporal domain (e.g., Botella, Arend, & Suero, 2004). In the latter situation, a series of stimuli is presented in rapid succession, and participants are asked to report the one stimulus in the sequence which has a different color or font. Errors in this task often involve the reporting of a stimulus in the dominant color or font, which means that the feature was attributed to the incorrect stimulus. Feature fusion resembles this situation. However, instead of binding the incorrect feature to one of the two objects, the features are combined and attributed to a single object. This does not only happen with sequences of two objects, but also with longer sequences, where it was shown that features can be transported across a stream of elements (Otto et al., 2006, 2009). In this sense, feature fusion and related paradigms address the problem of feature integration within a single domain. The feature fusion paradigm investigates possibly the most basic version of feature integration. The present results show that even in the extremely simplified case of feature fusion, complex processes take place that may only be captured by dynamical processes such as implemented in the Wilson–Cowan type model (Hermens et al., 2008) presented here. Apparently minor details of the stimuli, such as the almost invisible vernier offsets within grating masks, have clear and significant effects on participants’ performance on these stimuli by changing the spatio-temporal binding of the target vernier with elements in the successively presented grating. Investigating such low-complexity stimulus situations may be a first step in understanding how features across different domains, such as color and shape, are bound for stimuli presented across the entire visual field.

The model that we applied to explain our data provides a prediction for the conditions under which mis-attribute of features occurs. In particular, it suggests that binding in feature fusion is the consequence of the short presentation durations of the stimuli. For briefly presented stimuli, the interval until the activation in the excitatory layer is read out contains both the vernier and the anti-vernier. If the vernier and anti-vernier each would be presented for a longer duration (e.g., each for 100 ms), the signal related to the vernier and that to the anti-vernier could be read out separately, and feature fusion is not predicted to occur. Indeed, at longer presentation durations of the vernier and the anti-vernier and for larger offset sizes, fusion occurs less systematically. Instead, apparent motion or a superposition of the stimuli is often perceived (Scharnowski, Hermens, Kammer, et al., 2007). However, this seems to hold only for sequences that consist of two verniers. If instead a sequence of aligned and offset verniers is presented (e.g., Scharnowski, Hermens, & Herzog, 2007; Otto et al., 2006, 2009), fusion of vernier offsets appears to occur across a longer time interval. This suggests that either for objects that move across both space and time (Otto et al., 2006, 2009; Scharnowski, Hermens, Kammer, et al., 2007) or alternations of verniers and anti-verniers (Scharnowski, Hermens, & Herzog, 2007), more time is needed to close the object file (a representation of the currently attended object with its features) and more information can be integrated.

Our model suggests an important role for regular structures in visual perception. In this respect, it resembles models of texture segmentation (Thielscher & Neumann,
which also rely on the detection of edges and irregularities. As these models aim to describe early stage visual processing, which occur pre-attentively, the proposed mechanisms are in line with our findings of Experiment 1B. The implication would be, however, that vernier elements are considered as part of a texture rather than individual elements.

To conclude, feature fusion provides a versatile tool to study perceptual grouping over space and time. Our data suggest an important role for regular structures in such grouping operations. Simulations with a neural network model of our experimental data suggest that the detection of such structures is based largely on the information on the outside of these structures.

Acknowledgments

This work was supported by the Swiss National Science Foundation (SNF) project “The dynamics of feature integration.” FH is now a Postdoctoral Fellow of the Research Foundation-Flanders (FWO-Vlaanderen). We would like to thank Marc Repnow for technical support.

Commercial relationships: none.
Corresponding author: Frouke Hermens.
Email: frouke.hermens@gmail.com.
Address: Station 19, Lausanne, CH-1015, Switzerland.

References


