

Fig. S1. Photograph of a rat viewing a target with collinear flankers. The rat is trained to receive water from the port on the right when the central target is present, and the port on the left if it's absent.

Supplementary Data 1: Flanking stimuli impair target detection performance

The final step in our shaping sequence is the addition of flankers to the detection task (Fig. S2b). This made the task more difficult for rats, presumably for both cognitive and perceptual reasons. This study was not designed to explain this effect, which will be a constant across our compared conditions.

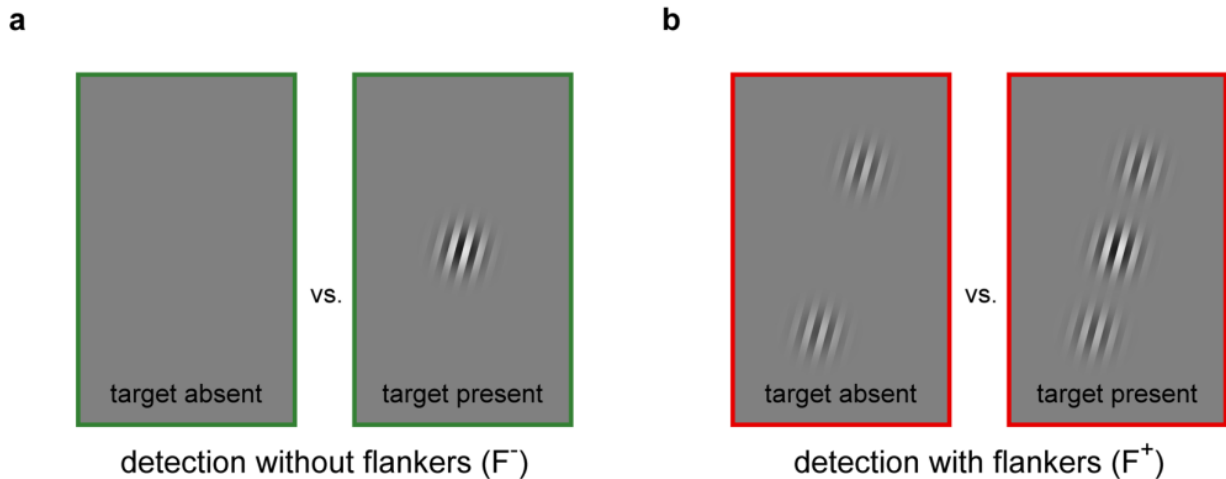


Fig. S2. Stimuli used for target detection, with or without flankers. a) During training (e.g. step 8), the rats perform a detection task when no flankers are present (F^-). When the rat initiates a trial, a target appears 50% of the time. When it is present, it is always located in the center of the screen. A rat would receive a water reward by correctly choosing to lick a sensor on the left side if the target were absent, or on the right side if the target were present. b) During the testing phase flankers are present (F^+) on 95% of the trials. The flanker's presence, location and orientation carry no information about the correct response. Only one flanker configuration is shown here, though all types were presented during testing (for more types see Fig. 4). In the first analysis all flanker conditions are grouped together.

To quantify the overall influence of flankers on rats' ability to detect a localized target, we compare detection with and without flankers, pooling over all other stimulus parameters (θ_T , θ_F , ω , S_T, S_F , see Fig. 4). During the continuous block of trials without flankers (Fig. S2a), rats performed about 75% of trials correctly. A single rat's performance is shown in Figure S3a, and for all seven rats in Figure S3b. Flankers impaired performance relative to both the previous block without flankers and the 5% of randomly interleaved trials in the testing step that omitted flankers. We note that performance on trials without flankers was lower when trials with flankers were interleaved as opposed to blocked, although the stimuli were identical. This implies that rats' visual processing or decision strategy depends on the distribution of stimuli over recent trials. For this and other reasons, all other results we show (except Supplementary Fig. S4c) compare only conditions randomly interleaved in the same block.

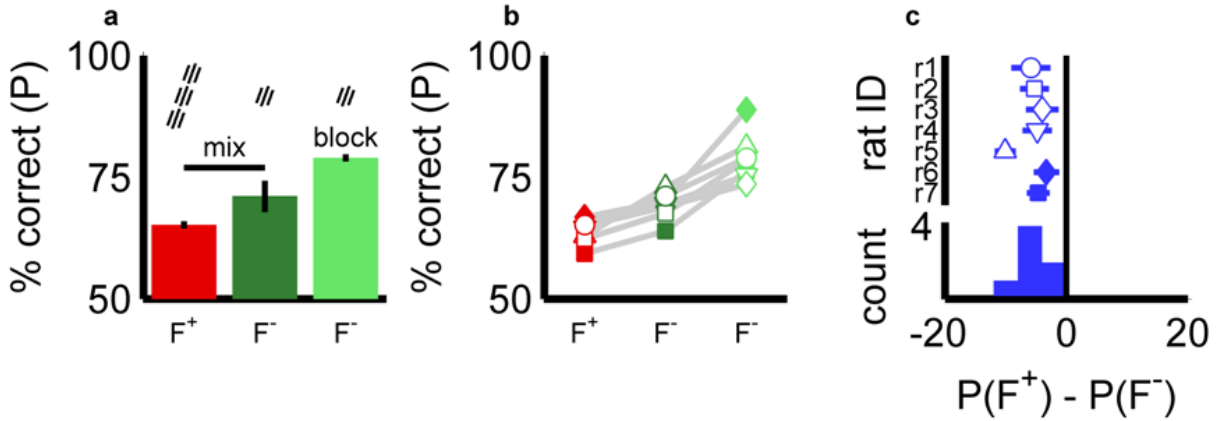


Fig. S3. Rats are worse at detecting the target when the flankers are present. a) The percent correct performance (P) of a single rat on three conditions with identical target properties. Error bars indicate 95% binomial confidence intervals. Performance from a block of randomly interleaved trials with flankers (F^+) and without flankers (F^-) is shown in red and dark green respectively. Light green indicates performance from a continuous block of trials without flankers (F^-) before flanking stimuli were introduced. b) Performance of seven rats in the same task, symbols colored as in (b), lines connect symbols representing a single subject (r1-r7). c) The effect size, measured as the difference in percent correct for interleaved trials with and without flankers: $P(F^+) - P(F^-)$. Symbols left of the zero line indicate that percent correct is lower when flankers are present. Effect is shown for each subject (symbol location indicates mean difference, horizontal lines indicate Agresti-Caffo 95% confidence interval, see Methods). For all individual subjects, the deficit with flankers is statistically significant. The effect is also significant if the difference between conditions is assessed at the population level ($p < 0.01$, Tukey-Kramer on 2-way ANOVA; $p < 0.01$, Tukey-Kramer on Friedman's test, see Methods).

The effect of flankers is summarized as the difference in performance on interleaved trials with and without flankers (Fig. S3c). If detection were independent of stimuli outside the target location, the difference would be zero. These data, however, show the difference to be less than zero, indicating worse performance with flankers.

We note that the addition of flankers also increases the probability that rats respond yes (Fig. S8a). This reflects a large increase in false alarms, and a small increase in the hit rate. We have not attempted to study or interpret this effect. These data are consistent with a modest reduction in “yes” responses due to contrast normalization, together with a large increase in “yes” responses due to task confusion or strategy, but our data by no means establish this interpretation.

We suspect that flankers confuse rats and also exert contrast normalization. Additional experiments would be required to isolate these components. Therefore we draw no conclusions about the underlying cause of the net effect of adding flankers, but see (Meier P, 2010). For this study, the effect is a constant across the compared conditions, and performance with flankers is sufficiently high that differences in performance between flanker configuration can be detected.

Supplementary Data 2. Choosing stimulus parameters

In order to test for influence of the surround, a detection task must be difficult enough to resolve a difference between conditions, but not so difficult that rats fall to chance performance. Difficulty is influenced by the target's size, contrast and spatial frequency as well as the proximity and contrast of the flanker. We trained rats on easy stimuli, and then varied one or two parameters at a time in separate experiments. These pilot experiments were done on different subsets of the subjects, as well as some subjects that were not in this study.

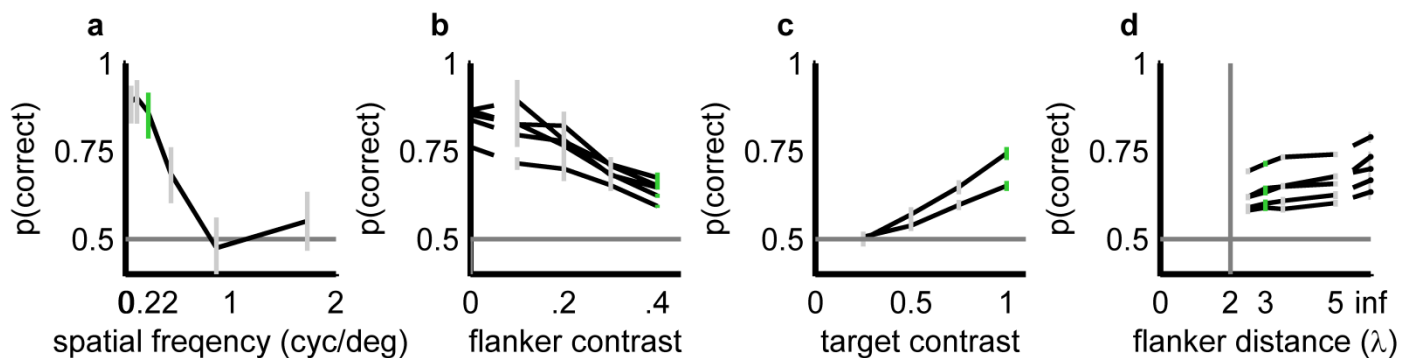


Fig. S4. Detection performance depends on target spatial frequency, target contrast, flanker contrast and flanker proximity. a) Spatial frequency dependence of a single rat's ability to detect a single large grating patch (Gaussian window of width 21.6 deg/std). Data were collected while randomly interleaving 5 contrasts and 5 spatial frequencies, but only data from 100% contrast (the target contrast used in testing) are shown. Performance was near chance for 0.86 or 1.73 cycles per degree (cyc/deg) at all contrasts, and was near maximum at 0.22cyc/deg at 100% contrast (green), which was used for all other tests in this paper for all rats. Error bars are the 95% binomial confidence interval. b) Detection performance for five rats while learning to detect targets in the presence of flankers of increasing contrast, using 0.22 cyc/deg gratings. Each rat is indicated by a separate curve. As rats demonstrated aptitude on lower flanker contrasts, they automatically graduated to the next higher contrast. All data points involve many hundreds of trials, with the exception of a single data point from one rat at 10% contrast involving only 75 trials. Performance for that rat may appear to be greater than performance with zero contrast flankers, but this difference is not significant. (Agresti-Caffo 95% confidence interval) All other tests in this paper used 40% flanker contrast (green). c) Detection performance from two rats with five randomly interleaved target contrasts, while flanker contrast remained at 40%. All other tests used a target contrast of 100% (green). d) Detection performance from five rats on stimuli with four randomly interleaved flanker distances. The line at $\lambda=2$ represents the distance that full contrast target and flankers begin to perceptually overlap for humans. At $\lambda=5$ both flankers are still on the screen, but are close to the edges. The data at flanker distance of infinity indicates trials where there was no flanker on the screen. All other data in this paper uses $\lambda=3$ (green).

We tested spatial frequencies spanning 5 octaves (0.05-0.86 cyc/deg) and five contrasts (12.5-100%), for a total of 25 interleaved conditions. This test was performed using a large target with no flankers present (see Fig. 1d, step 7 for example stimulus). At 100% contrast, rats performed sufficiently well on 0.22 cyc/deg (green, Fig. S3a), but not well enough for 0.43 or any higher spatial frequencies. This is consistent with the visual contrast sensitivity previously reported for Long Evans rats; 0.22 cyc/deg is well within their capacity, but a little higher than their peak sensitivity. Gratings were 0.22 cyc/deg for all stimuli presented during the testing step in the main results, as well as all of the other sub-panels in Figure S4.

When flankers were present, they induced a deficit in performance that increased with their contrast (Fig. S4c). This sub panel is the only data in this paper (other than Fig. S3a,b) that was not randomly interleaved between conditions, but was derived from training steps in which the rats were automatically shaped to higher contrast flankers (see Fig 1d, step 9 for example stimulus). We know that the rat's behavior during this learning stage was not constant over time. In principle, learning effects could have increased a rat's performance at higher contrasts which were tested at a later date. Yet flanker impairment overwhelms any such learning improvements. We include this panel because it is informative about the rats learning and about the difficulty of the detection task in the presence of flankers. The trend is the same for other rats in which a range of flanker contrasts were revisited after learning (data not shown).

After rats were well trained on detecting targets in the presence of flankers, we examined the influence of target contrast. Interestingly, rats were sensitive to the contrast of the target across the entire range of our linearized monitor. If stimuli are supra-threshold on a detection task, one would expect performance to plateau. However, if increasing contrast improves performance, this is consistent with the target stimulus being at or near threshold for detection. Since contrast limited performance, even at the upper end of our display, we chose a high contrast target, so that we could better resolved differences in performance (near 75% correct). We believe that the threshold for detection is shifted to a higher contrast by virtue of the flankers in the surround.

Flankers have a greater influence on target detection when they are closer. We note the possibility that subjects would be best when flankers are at a distance of 3λ , which they had prior training with. However, we observe rats perform worse at 2.5λ and better at 3.5λ . Despite the stronger effect at 2.5λ , our main results use a flanker distance of 3λ because previous work in human psychophysics and cat V1 neurophysiology find effects at this distance. Additionally, the tails of the Gaussian masks for target and flanker begin to overlap and 2.5λ , even though we truncate tails at 4 std. There is no overlap for our flankers and targets at a distance of 3λ .

Taken together these data suggest that the flanker stimuli used in our study are perceptually difficult for rats, because the spatial frequency is high for a rat, the grating patch small, and the flanker contrast high. The sensitivity to changes in contrast of the target or the contrast or distance of the flankers is all consistent with expectations from contrast normalization. Due to confounding cognitive factors we do not purport to measure the strength of contrast normalization from these data.

Supplementary Data 3: Statistics for all pairwise comparisons of flanker configurations

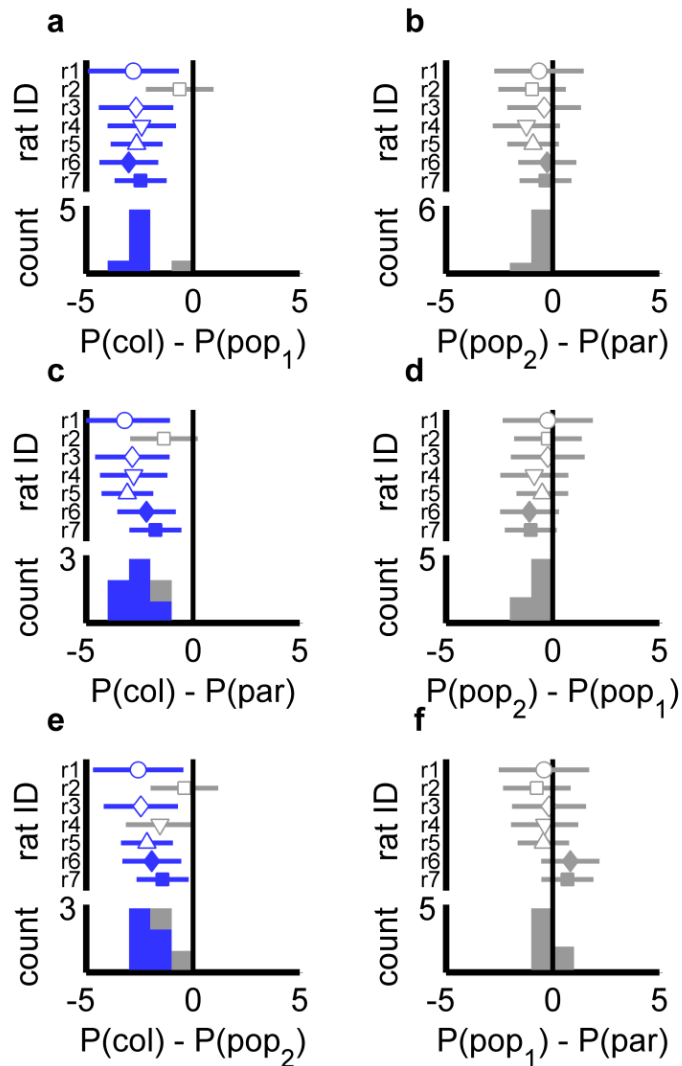


Fig. S5. Additional pair-wise comparisons between the flanker conditions. We present all six comparisons between the four stimulus conditions measured by the difference in percent correct. Panels a and c are the same as Figure 4c,d; they are reproduced here to facilitate comparisons. The vertical midline indicates that the performance on the two conditions is not different. Each horizontal line represents the 95% confidence interval for a single subject's data on a given comparison (Agresti-Caffo). The top row of sub panels isolates comparisons between pairs of conditions in which the location of the flanker differs, but other parameters were held constant: a) collinear – parallel and b) popout₂ – popout₁. The middle row isolates comparisons in which only flanker orientation differs: c) collinear – popout₁ and d) popout₂ – parallel. The bottom row captures comparisons in which only target orientation differs: e) collinear – popout₂ and f) parallel – popout₁. The three comparisons that contain subjects with significant effects (Agresti-Caffo 95% confidence interval) are best explained by the reduction of performance for collinear stimuli which affects all the comparisons in the left column. The comparisons in the right column are not significant for the rats individually, nor at the population level (see Fig. S6).

In the main paper we show statistics for comparisons between the collinear condition and two others: the condition that differs only in flanker orientation (Fig. 4c) and the condition that differs only in the flankers' angular position (Fig 4d). We consider all pairwise comparisons here.

For the four stimulus conditions tested, there are a total of six pair-wise comparisons. Rats performed worse on collinear stimuli than on any of the other three stimulus conditions (Fig. S5a,c,e, left column). Each of these comparisons is the result of disrupting collinearity in a different way: by changing the flanker orientation (a, same as Fig. 4c), by changing the flankers' angular position (c, same as Fig. 4d), or by changing the target orientation (e). Taken alone, it seems possible that the visual system may be sensitive to each of these changes in isolation. However, the other three conditions provide a control (Fig. S5b,d,f; right column); in each the geometry of the stimulus was held constant except for the flanker orientation (b), flanker location (d), or target orientation (f). No individual rats are significant in these three control comparisons (Agresti-Caffo 95% confidence interval), and they do not differ at the population level (Tukey's on ANOVA, $p > 0.05$; Tukey's on Friedman's test, $p > 0.05$).

To communicate significant differences at the population level, we present the marginal mean performance for all rats (Fig. S6). Each error bar represents the half width of the critical value from Tukey's test of Honestly Significant Difference, which accounts for multiple comparisons. If the error bars from two conditions do not overlap vertically, then the difference between conditions is significant. The figure shows significance for $p < 0.05$; we also assessed significance at $p < 0.01$. All multiple comparisons in this paper use this same method; here we provide a graphical view to facilitate understanding. Figure S6a summarizes the results for the ANOVA, and Figure S6b summarizes Friedman's test.

In one instance (collinear vs. popout₂, Fig. S6), the difference is significant at the population level by 2-way ANOVA but not by Friedman's test after adjusting for multiple comparisons. (Both tests on this condition are significant at $p < 0.01$ before adjusting for multiple comparisons). Friedman's test is non-parametric and only acts on the rank-ordered average performance. With only seven subjects, it is quite conservative. It is actually surprising that two of the six comparisons are significant by this test even after adjusting for multiple comparisons. We interpret the collinear condition to be harder for rats than the popout₂ condition, based on the results of the 2-way ANOVA and the facts that this was true for 7 of 7 rats, and significant for 5 of 7 rats.

We find it somewhat surprising that there is no discernable effect of popout with respect to the parallel condition (Fig. S5b,f). One might have expected popout conditions to be easier. Maybe the presence of only two flankers does not indicate the uniqueness of the center target as much as a field of distracters would (Nothdurft, 1991; Smith et al., 2007). Alternatively, the orientation difference between target and surround may be important. We used 30° (open symbols in Fig. S3) and 45° (filled symbols), and did not observe a difference. A popout effect might be found at more orthogonal angles (Nothdurft, 1991; Schwartz et al., 2009).

In summary, our only significant differences involved comparisons to the collinear condition. Thus, impairment on the collinear stimulus is the primary effect of pattern-specific processing that we observed.

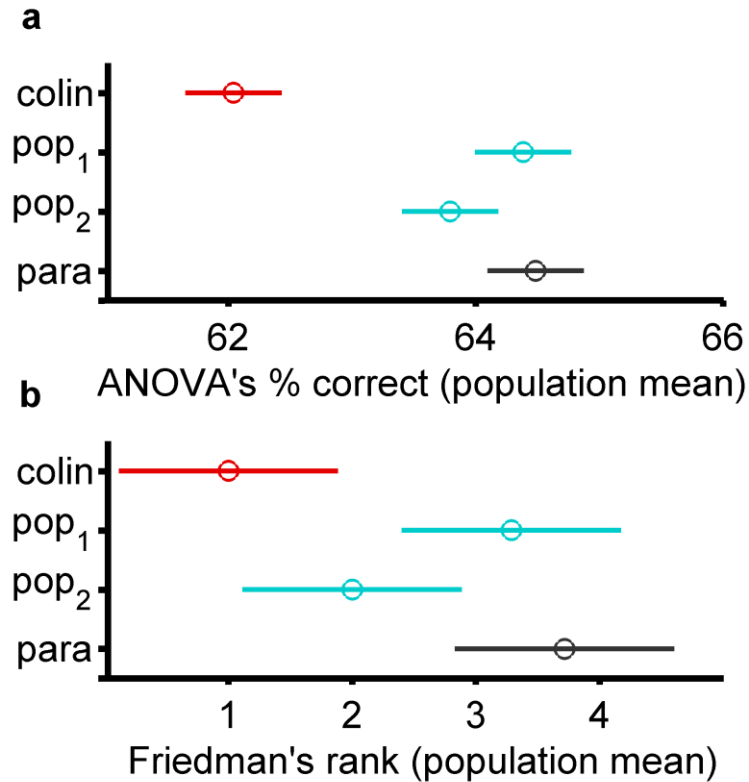


Fig. S6. Multiple comparison adjustments. a) A visual summary the 2-way ANOVA after having adjusted for multiple comparisons. The x-axis is the marginal means of the percent correct performance from a population of seven rats. The error bars represent the half width of the critical value for multiple comparisons using Tukey's Honestly Significant Difference at $p < 0.05$. If the tails of two error bars do not overlap each other on the vertical axis, then the difference those two conditions are significantly different at $p < 0.05$. b) A visual summary the Friedman's test after having adjusted for multiple comparisons using the same method. Friedman's is a non-parametric test that does not assume Gaussianity; it tests for a shift in the location of a probability distribution by analyzing the relative rank of performance across rats and conditions. Friedman's has less power to reveal an effect, especially with the small $N=7$. The collinear condition is reliably the 1st rank (worst performance) for all rats. The popout₁ and parallel conditions have higher performance. Interestingly, popout₂ is reliably the 2nd rank, but this test could not resolve a significant difference from the collinear condition. We do not interpret this strongly because the absence of observing an effect is inconclusive, and the population marginal means shown in a), suggests more strongly that popout₂ groups with popout₁ and parallel more than with collinear. Additionally, the significance of the each of the rats individually (Fig. S5) supports the conclusions of the 2-way ANOVA. All tests that are significant in this figure (both a and b) are also significant at $p < 0.01$ (not shown).

Supplementary Data 4: No influence of phase

In our experiment, for each flanker orientation and position we also varied sign of the grating ($S_F = \pm 1$) corresponding to inverting the luminance of the grating such that dark bars are switched for light bars. Changing the sign is equivalent to a π shift in the spatial phase of the grating. Neither the absolute sign of the target or flanker nor the relative sign between them had an effect on performance. Here we show there is no difference between the phase aligned and phase reversed collinear stimuli using the same data from the seven rats reported in the main paper (Fig. S7a,c). This experiment used one spatial frequency (0.22 cyc/deg), one flanker distance (3λ , 13.6°), and two relative phases (0, π). Our data do not exclude the possibility that the rat visual system is sensitive to phase differences with other choices of these parameters.

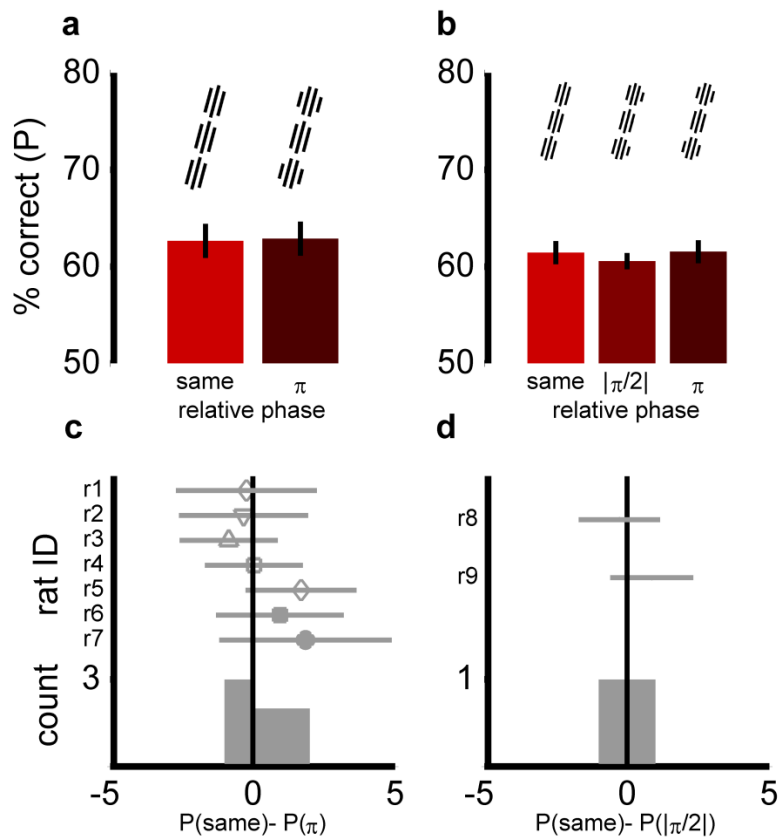


Fig. S7. Phase of grating has no influence on detection with flankers. a) A single rat's performance on collinear stimuli that were either phase-aligned or shifted by π (sign-reversed). This is the same data as Fig. 4, reanalyzed to show that there is no effect of phase. b) A single rat in a separate pilot study in which four target phases and four flanker phases were used. Conditions are grouped by the phase difference between target and flanker. c) The difference in percent correct between phase-aligned and phase-reversed trials for the seven rats in the main study. No individual rats show significant differences. The population as a whole shows no indication that the relative phase of the flankers influences the rats detection performance ($p > 0.05$, ANOVA; $p > 0.05$, Friedman's test). d) The same is true for the pilot study with more phases. All three pair-wise comparisons are insignificant for each individual rat ($p < 0.05$, Agresti-Caffo interval). Shown is the difference between phase aligned and stimuli in which the phase was shift $\pm\pi/2$. No population statistic was performed because $N=2$.

In a pilot study (N=2) we tested four different absolute grating phases for target and flanker, using the same class of flanker stimuli as this study, but with some differences: the target contrast was weaker (60% instead of 100%) and rat performed a two alternative force choice between two simultaneously presented stimuli, instead of a single stimulus yes-no task. The relative phase between target and flanker did not influence the detection of the target (Fig. S7b,d).

Since there is no evidence in our data of sensitivity to sign or phase, we have no reason to include it as a relevant parameter in the schematic model presented in the main paper (Fig. 6). The model is sufficient to explain all of the results in this paper, without adding a term in the surround for the relative phase.

The influence of oriented contrast in the surround is not phase sensitive in many previous psychophysical(Xing and Heeger, 2001) and physiological experiments(Webb et al., 2005). However, sometimes the relative phase of the surround does play a role(Ejima and Takahashi, 1985; Williams and Hess, 1998). For primates performing contour detection tasks, spatial pattern summation is phase insensitive in the fovea, but in the periphery it is not(Chen and Tyler, 1999). While phase sensitivity attests to the specificity of surround processing, it has been argued that invariance to phase is a useful aspect of surround processing, and such invariance continues to be applied in computational models of contour integration(Hansen and Neumann, 2008).

Supplemental Experimental Procedures 1: Water restriction, training schedule and environment

Beginning around post-natal day 30, rats were restricted from free access to water, instead working for water reward in the training environment. Water was earned solely by correctly performing tasks (“closed economy” training) as long as animals maintained adequate health and hydration. This was assessed by daily weight and health inspections. Supplemental water (or hydrating snacks, e.g. carrot slices) were provided as needed, for less than 25% of rats during initial shaping, and rarely after 2 weeks. On most days, including weekends, animals were transferred to the training chamber where they could freely perform trials for around 90 minutes. On days where sessions were skipped, rats were provided with about an hour of free access to water, but we avoided this because it reduced their motivation to perform trials on the following day. Continuous access to rat chow was provided both in their home cages and the training chamber to stimulate desire for hydration. We phase-reversed the light cycle in the room so that rats trained during the day in dark rooms illuminated almost exclusively by the glow of 7 computer displays. When overhead lighting was required for working in the room during daytime, we typically filtered out the high wavelengths visible to rats(Jacobs et al., 2001) and worked in red light (Encapsulite Intl., red filter, 625nm cutoff, 48SOR20T12).

Supplemental Experimental Procedures 2: Training system

The system consists of an array of stations, each controlled by a computer running Windows XP-Pro and Matlab. An additional computer is used as a single point of control for all the stations (via custom TCP/IP code in Java/Matlab). This allows sessions to be started/stopped en masse, management of subject information and trial records in coordination with a database (Oracle 10g Express Edition), and centralized management of individualized training sequences and parameters for each subject.

The system design facilitates either live-in training or easy swap-in of groups of animals. The animals in this study were members of one of 6 groups that were swapped in daily. Since rats can earn hydration adequate for health in roughly one hour of trials, we found it efficient to limit each animal's daily access to trials to sessions of about this duration. This focuses their motivation for performing trials correctly throughout the session and multiplies the number of subjects that can be trained using a given amount of available hardware.

Each training box is 35 cm wide x 18 cm deep x 30 cm high, with detector/reward ports positioned along the front wall spaced 9 cm apart 6 cm above the floor grating. The detector housings protrude 2.5 cm into the box, and the CRT is positioned 5 cm behind their wall. At the center port, a rat's eye is roughly 10 cm from the monitor, 10 cm below its center. At this position, the monitor display subtends 104° of visual angle, and the target grating at the CRT center is roughly 14 cm away and spans 6.8° per std of the Gaussian mask. This is closer than the distance established for maximum behavioral visual acuity in rats (20-30 cm; Wiesenfeld, Branchek 1976).

Accurate reward delivery requires a system that does not store pressure in unexpected ways. At such small volumes and flow times, tiny differences in port geometry likely affect the amount of reward and its accessibility due to being wicked away or the force with which it is ejected, and reward volumes/utility may not be linear with respect to open valve duration. This can cause strong side biases. We monitored behavioral trends for side bias and found we could control it using timeouts and correction trials alone; we only sporadically and imprecisely verified that left and right reward drops looked roughly similar in size. We have since added a syringe-pump based reward system that is more accurate and requires less maintenance (New Era Pump Systems, Inc. NE-500).

We used the parallel port for electronic interface of the valves and sensors with the computer, using PortTalk (Craig Peacock, <http://www.beyondlogic.org/porttalk/porttalk.htm>) via an open source Matlab wrapper (<http://psychtoolbox.org/wikka.php?wakka=FaqTTLTrigger>).

The sound for requests is a tone with energy at each octave spanning the frequency range of the speakers (some range within 20-20,000 Hz); for correct responses, a tone with the same harmonic structure is played a perfect fourth above the request tone, creating a harmonic resolution. The sound for incorrect responses is a

chord made from two tones of the same harmonic structure separated by a tritone, the maximally dissonant interval. The sound for inappropriate responses is broadband noise. Sometimes a drop of water lodged in a port is sufficient to break the infrared detector beam, in which case, the corresponding sound plays continuously until the water is cleared by the rat (after some experience in the box, rats are observed to clear ports in this way even during a period of no trial activity, indicating a possible preference for silence to the continuous sounds). No provisions were made for sound isolation. Sounds from adjacent boxes are quite audible in any given box, but quite attenuated relative to that box's local sounds. The CRTs for adjacent boxes are not visible when operating the ports because stations are separated by opaque dividers.

Our software architecture facilitates the design of arbitrary task structures and manages each subjects' progress through their own training protocol. A *protocol* specifies a sequence of *training steps*, each with a *trial manager* (defining task structure – e.g. two-alternative-forced-choice), *stimulus manager* (defining audiovisual stimuli and their parameters), *reinforcement manager* (specifying reward/timeout reinforcement rules), and *graduation criteria* (regulating progression through steps).

Supplemental Experimental Procedures 3: Stimulus display

We used CRT displays (NEC FE992, 19") because of their fast refresh and phosphor decay times; CRTs are generally better suited for visual psychophysics than LCD flat panels because their timing artifacts and brightness artifacts are more consistent and better understood. The CRT linearization table was created by fitting a power law with gain and offset ($y=b*x^\gamma+m$) to photodiode measurements (Thorlabs, PDA55), and then computing the inverse function ($x=[(y-m)/b]^{1/\gamma}$) for each RGB channel independently. Each value was measured as the average height of the smoothed peak of the phosphor decay curve recorded with that value presented in a rectangle occupying the central 60% of screen for 9 frames at 100 Hz and 1024x768 resolution. Gun values were measured in increasing order rather than randomized. The resulting tables were then verified to have linearized grayscale output to within 0.5% using the same method.

Since the background grey level was set equal to the mean luminance of the grating, linearizing has the consequence that a blurry optics system with spatial resolution lower than the spatial frequency of the gratings will not perceive a luminance difference at any point in the Gaussian patch of the grating, but only a smooth grey at the same luminance as the background. We confirmed that this was the case for sufficiently high spatial frequencies for human observers. The linearization range was chosen such that the gratings were effectively lower contrast than in previous steps and the mean grating luminance/background grey was brighter than in previous steps. The minimum, mean, and maximum luminance were set to 4, 42 and 80 cd/m^2 , respectively (Colorvision, spyder2express).

Table S1: Training Details

Step	Description	Visual Stimulus	Graduation Criterion	Goal	Duration (trials \pm std)	Duration (days \pm std)
1	Free drinks	none (mean gray)	4 trials in 1 min	get water from port	100 \pm 65	1 \pm 0
2	Earned drinks	none (mean gray)	5 trials/min for 2 min	alternate responses	170 \pm 100	2 \pm 2
3	Faster drinks	none (mean gray)	6 trials/min for 3 min	sustain interest	170 \pm 100	2 \pm 2
4	Side rewards	big grating	5 trials/min for 5 min	2AFC trial structure	350 \pm 220	8 \pm 5
5	Easy detection	same big grating	85% correct	1 st visual learning	4,100 \pm 3,300	17 \pm 13
6	Linearized luminance	lower contrast	85% correct	new stim	1,700 \pm 1,300	9 \pm 9
7	Thinner	double target cyc/deg	85% correct	new stim	900 \pm 1,100	4 \pm 3
8	Smaller	target mask 1/3 rd size	85% correct	new stim	7,300 \pm 8,200	35 \pm 45
9	Add flanks	slowly raise flank contrast	80% correct	new stim	11,000 \pm 9,000	25 \pm 13
Test	Flankers	See Fig. 4b	none	collect data	55,000 \pm 23,000	98 \pm 36

Table 1. Shaping sequence and training steps. Step numbers correspond to the numbers in the colored bars in the Figure 1c training timeline. All rats progressed from step 1 to 9 in increasing order. Two rats (r1, r2) performed contrast and spatial frequency varying psychometric curves before step 8 (not listed in this chart, results in Fig. S4a). Another two rats (r4, r8) performed the flanker task with varying target contrast (not listed in this chart, Fig. S4c). Other supplementary tests either used data from the training sequence (flanker contrast, step 9, Fig. S4b), the main testing step (influence of luminance sign, Fig. S7a,c), or after the main testing step (flanker distance, Fig. S4d).

Supplementary Analysis 1: Considering signal detection theory

In the main paper we summarize performance as percent correct, but using d' does not alter the character of the results (Fig. S8). Signal detection theory measures attempt to unconfound bias and sensitivity, both of which affect percent correct. The metric of percent correct can be misleading when comparing results of subjects with very different overall biases. In this study we do not directly compare any rat's absolute performance to another rat's. Rather we test within subject for significant differences between stimulus conditions that are randomly interleaved. When the hit or false alarm rate is near 0 or 1, it is possible that the metrics of d' and percent correct could give different trends or statistical significance when comparing

conditions. In our data, the hit and false alarm rates are bounded by [0.15 0.85] for all rats and conditions. Thus our data do not occupy the extreme region where one would expect significant differences between the metrics.

One advantage of using percent correct, when the data permit, is that it avoids the need to make assumptions which we cannot verify. Specifically, signal detection theory separates bias from sensitivity by assuming normality and homoscedacity of the stimuli distributions being distinguished – d' summarizes their discriminability as the separation of their means in units of their common standard deviation. Another reason we prefer percent correct is that, unlike d' , it monotonically increases with reward; thus, tracking what rats naturally prioritize. We report the assumption-free data on hit rate and false alarm rate from which d' and other metrics of signal detection theory are easily calculated.

Nevertheless comparisons based on d' are provided for comparison (Fig. S8). Across every major comparison in this study, the two metrics agree in individual significance, population trends, and relative magnitude. Specifically, we find that both measures indicate all flanker types impoverish performance (Fig. S6b,c), collinear flankers impair performance more than parallel flankers (Fig. S6 e,f), and collinear flankers maintain this impairment regardless of the relative contrast sign between target and flanker (Fig. S6 h,i).

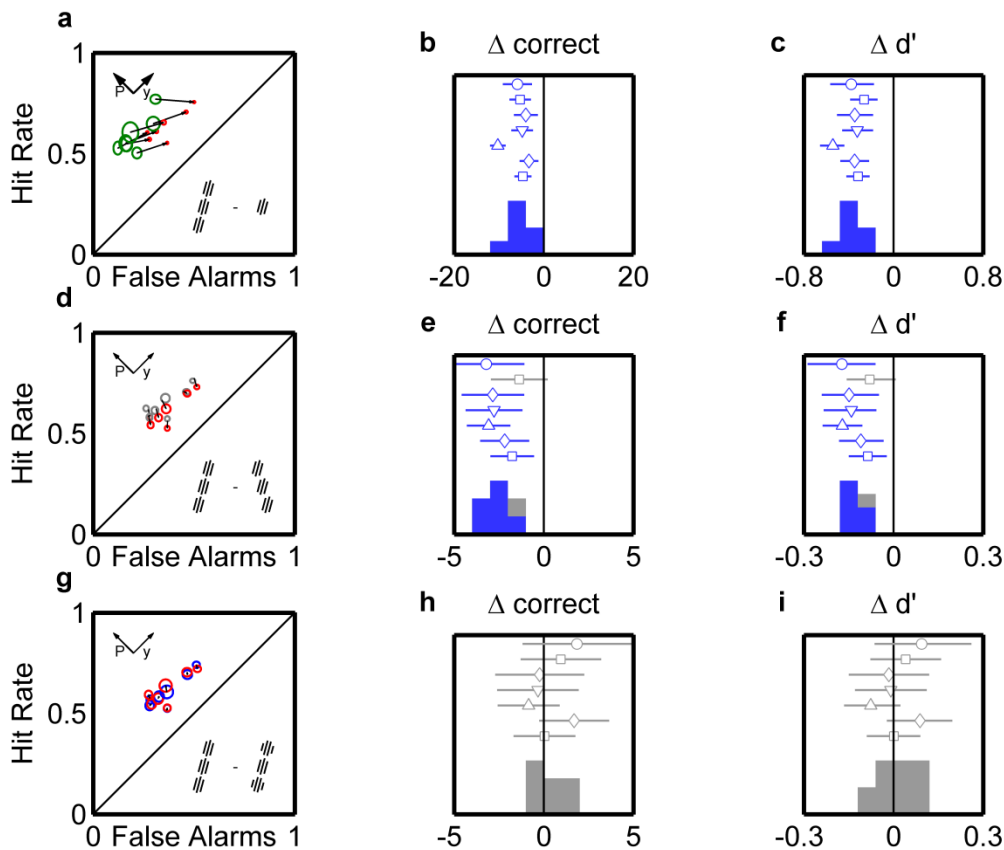


Fig. S8. Percent correct and d' yield the same conclusions. The raw data are presented for three experiments: the influence of flankers (a,b,c), the influence of collinearity (d,e,f), and the influence of phase (h,i,j). The purpose is to validate that the conclusions are the same using percent correct or d' as a performance metric. a) Same data as in Figure S3, comparing the difference between two conditions summarized by the icons lower right of the panel. The ellipses indicate performance of each rat on interleaved trials with the flanker present (red) or absent (green). Arrows indicates a single subject's change in

hit rate and false alarm rate. The axis of hit rate and false alarm can be rotated to indicate percent correct performance (P) and percent yes responses (Y), as indicated by the inset axis in the upper left. All rats have a decrease in the percent correct when flankers are present. *b*) Performance reduction quantified by percent correct using Agresti-Caffo 95% confidence interval (identical to Fig. S3c). Each horizontal line is the confidence interval for a single rat. The histogram of the mean change in performance color coded blue for each rat that is individually significant, and grey otherwise. *c*) Performance reduction quantified for d' using the 95% confidence interval of an MCMC simulation. Panels (*d,e,f*) employ the same convention as (*a,b,c*), except comparing collinear flanker condition (red) to the parallel flanker condition (gray). Data are the same as Figure 5. *e*) difference by percent correct metric (identical to Fig. 5d). *f*) difference by d' metric (identical to Fig. 5e). Panels (*h,i,j*) show the lack of an observed difference between collinear stimuli that were phase aligned (red) or phase reversed (blue). The data are the same as Figure S7 a,c. The effects in panels *b*, *c*, *e* and *f* were all significant by ANOVA ($p < 0.01$) and by Friedman's test ($p < 0.01$), adjusted for multiple comparisons where required. There was no significant difference in phase (panels *h* and *i*) at the population level ($p > 0.5$ both ANOVA or Friedman's test).

Supplemental References

- Chen, C.C., and Tyler, C.W. (1999). Spatial pattern summation is phase-insensitive in the fovea but not in the periphery. *Spat Vis* 12, 267-285.
- Ejima, Y., and Takahashi, S. (1985). Apparent contrast of a sinusoidal grating in the simultaneous presence of peripheral gratings. *Vision Res* 25, 1223-1232.
- Hansen, T., and Neumann, H. (2008). A recurrent model of contour integration in primary visual cortex. *J Vis* 8, 8 1-25.
- Jacobs, G.H., Fenwick, J.A., and Williams, G.A. (2001). Cone-based vision of rats for ultraviolet and visible lights. *J Exp Biol* 204, 2439-2446.
- Meier P, R.P. (2010). Model of visual target detection applied to rat behavior. *Front. Neurosci. Conference Abstract: Computational and Systems Neuroscience 2010*.
- Nothdurft, H.C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Res* 31, 1073-1078.
- Schwartz, O., Sejnowski, T.J., and Dayan, P. (2009). Perceptual organization in the tilt illusion. *J Vis* 9, 19 11-20.
- Smith, M.A., Kelly, R.C., and Lee, T.S. (2007). Dynamics of response to perceptual pop-out stimuli in macaque V1. *J Neurophysiol* 98, 3436-3449.
- Webb, B.S., Dhruv, N.T., Solomon, S.G., Tailby, C., and Lennie, P. (2005). Early and late mechanisms of surround suppression in striate cortex of macaque. *J Neurosci* 25, 11666-11675.
- Williams, C.B., and Hess, R.F. (1998). Relationship between facilitation at threshold and suprathreshold contour integration. *J Opt Soc Am A Opt Image Sci Vis* 15, 2046-2051.
- Xing, J., and Heeger, D.J. (2001). Measurement and modeling of center-surround suppression and enhancement. *Vision Res* 41, 571-583.