The Limits of Counting

Jake Baron

Stuyvesant High School

Professor Denis Pelli, NYU, Mentor

Anne Manwell, Adviser

J. Baron (2006) The limits of counting. Intel Science Talent Search.
http://psych.nyu.edu/pelli/highschool.html#2006

The Limits of Counting

Abstract

Many everyday tasks require counting. It is hard to imagine a situation in which we might be unable to count a small group of objects before us. But we find here that counting ability is much more prone to error than it may seem. Even under the best conditions, we can accurately count only up to 3 objects in one glimpse. Moreover, when counting objects out of the corner of the eye, performance is drastically impaired if the objects we wish to count are near other objects we wish to ignore. Under this condition, the standard deviation of observers' counts of up to 3 objects is about 1.5, indicating surprisingly great error for so few objects. Our results indicate that accurate counting of cluttered objects cannot be done in a peripheral glimpse. Each of the objects must be fixated individually, one at a time.

Introduction

In 1871, William Stanley Jevons made an important finding in visual psychophysics: one's error in estimating a number of objects in the visual field is directly proportional to the number of objects one attempts to estimate. Though this result has been replicated, little further progress has been made in exploring visual counting. For instance, until now, there has been no investigation of our ability to count objects viewed in the *peripheral* field of vision (that is, everything one sees except the small area at which one looks directly, or what is commonly called the corner of the eye). Here we ask:

1. Can we count a group of objects as well in our peripheral vision as we can in central vision?
2. If so, is this true in all cases, even when the visual scene around the objects is complex?
3. If not, how great is the impairment in our ability? And finally,

4. Knowing this, how can we avoid counting errors in real life?

We sought to investigate counting ability in peripheral vision as compared to that in central vision. It was therefore necessary for some of the objects we would present to appear in central vision and others to appear in peripheral vision. For practicality, the “objects” we used in testing were small black dots presented on a computer screen. To determine the dot presentation time we would use for our testing, we reasoned that if our observers had no time limit for counting the dots, they would count very carefully by *fixating* (looking directly at) each one individually, and thus they would report the correct number of dots every time. For this reason we presented the dots briefly (200 ms), allowing only one glimpse. (We wanted to study the ability to count objects seen in peripheral vision, so obviously we couldn’t have our observers fixating the objects.) 200 ms is enough time to see the dots, but not enough to move the eyes to shift fixation onto the dots (see Carpenter, 1988).

Methods

Design. Observers were put to three counting tasks, each repeated numerous times; each repetition was termed a trial. The independent variable in all three tasks was the number of dots presented that the observers were directed to count. This was termed *numerosity*, following convention in the field (for example, Feigenson, Dehaene, and Spelke, 2004). For practicality, we restricted our investigation to the numerosities 1 through 30. The dependent variable in all three tasks was *error*, defined specifically for the purposes of this study as the standard deviation of an observer’s numerical responses when asked to count dots at a given numerosity in a given task. For example, if an observer counted the following numbers of dots at numerosity six of task 2, {6, 6, 7, 4, 6, 6, 6, 5, 6, 6, 6, 8, 6, 4, 5, 6, 6, 6, 6, 6, 6, 6, 6, 7, 6, 7, 6, 6, 6, 4}, then his *error* for numerosity six of task 2 would be 0.845, the standard deviation of these numbers. Standard deviation, as opposed to root-mean-square error or another measure of absolute deviation, was used to measure inaccuracy for the following reason: inaccuracy has two components, mean error and standard deviation. The mean error varies between individuals initially, but it approximates zero in practiced observers (Jevons, 1871). Still, standard deviation represents a more stable and fundamental source of inaccuracy. We have made plots, not shown, of the root-mean-square error, but they are negligibly different from the shown plots of standard deviation.

The observers' three tasks were designed to answer the motivating questions presented above. Task 1 asked the observers to fixate the center of a shown rectangle and to estimate the number of dots briefly presented within it. (The rectangle was called a "box.") This central-viewing task served as a control for task 2, which was identical to task 1 except that observers fixated a specific designated point above the rectangle. Task 2 was in turn a control for task 3, which was identical to task 2 except that we added "distracter" dots outside the box—observers were directed *not* to include these dots in their counts. (For all three tasks, the basic direction was "count the number of dots in the box.") In this way, tasks 1 and 2 answered our questions regarding the degree of impairment in counting ability, if any, when the objects are viewed peripherally as opposed to centrally. Tasks 2 and 3 answered our question regarding whether we can count in the periphery as well when the scene is crowded as we can when it is not.

Participants. Five observers were tested in this study. They gave written consent before testing began, after the procedures had been explained to them, and they were paid for their participation. Their ages were 10, 16, 17, 35, and 58 years (mean age = 27.2 years). Three were female and two were male, and all had normal or corrected-to-normal visual acuity.

Apparatus. The computer monitor on which the test stimuli were shown was a Dell Trinitron 16-inch CRT monitor with frame rate 100 Hz and resolution 1024 x 768 pixels (width-byheight), whose screen measured 30.5 x 22.9 cm. We wrote the computer programs to present our test stimuli in the Matlab 7 programming language from The Mathworks, with the help of the Psychophysics Toolbox extension (Pelli, 1997; Brainard, 1997).

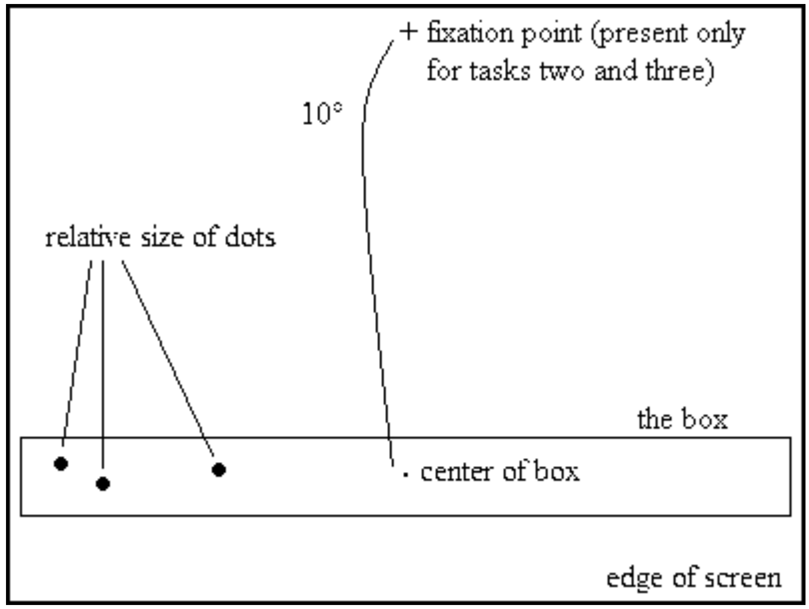

Figure 1. Schematic diagram of the stimulus screen presentations. The box was presented in the same place on the screen in all three tasks. In task 1, observers fixated the center of the box directly, while in tasks 2 and 3 they fixated the plus sign at the top of the screen, which is ten degrees above the center of the box. The stimulus background was white, and the dots, box, and fixation point were black, as shown. The dot diameter was 0.4°, and the thickness of the edges forming the box was 0.02°.

Figure 1 is a diagram of the basic appearance of the test stimuli that observers were shown. All the dots to be counted always appeared in the box, not touching the edges of the box, and not overlapping each other. The separation between the center of the box and the fixation point for tasks 2 and 3 was ten visual degrees when viewed from a perpendicular distance of 75 cm, which was the universal viewing distance in this experiment. The diameter of the dots was a constant 0.4°, and the minimum distance between the dots was 0.4° edge-to-edge. The computer programs placed the dots randomly for each stimulus screen independent of their placement for all the other screens. Finally, for each task 3 stimulus screen, 90 distracter dots were presented in the regions immediately above and below the box; 45 in each region. These dots were identical in every way to the dots in the box, and their spacing and random manner of placement were the same as well. Figure 2 shows examples of stimulus screens from the three tasks.

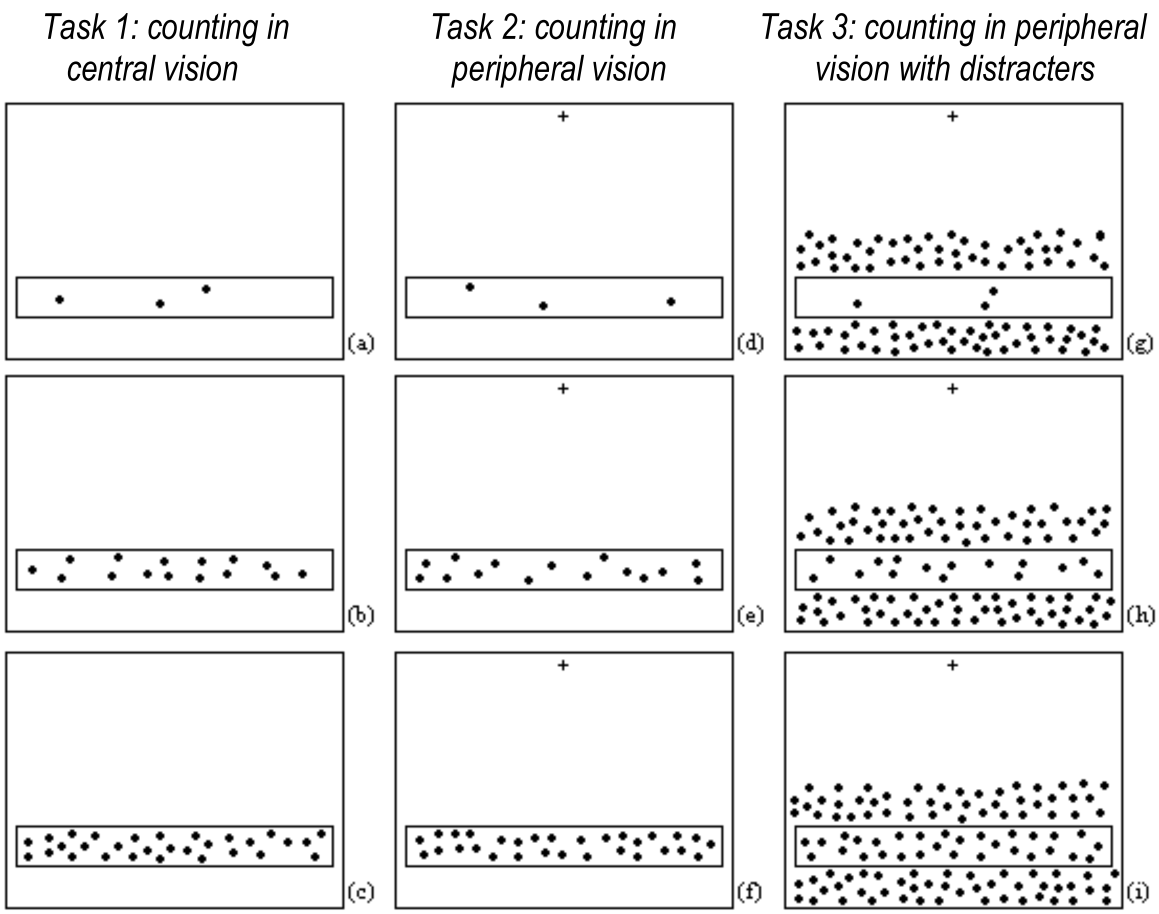

Figure 2. Panels (a), (b), and (c) are examples of task 1 stimulus screens of numerosities 3, 15, and 30, respectively. Panels (d), (e), and (f) are the same for task 2, and panels (g), (h), and (i) are the same for task 3.

Procedure. Each observer responded to 30 stimulus screens of each numerosity, one through 30, for each task, so each completed a total of 2,700 trials: 30 screens per numerosity x 30 numerosities per task x 3 tasks. These were divided into nine 300-screen sittings, with three sittings entirely devoted to stimulus screens of task 1 and six sittings containing a random mix of stimulus screens from tasks 2 and 3. In both cases, the computer programs randomized the screen presentation order with respect to numerosity, and in the latter case, the order was randomized with respect to task as well. The reason that stimulus screens of task 1 were isolated by sitting from those of tasks 2 and 3 was to ensure that the fixation point of all screens in a given sitting was the same. This was done for the important practical reason that the experiment would have taken much longer otherwise, because the observers would have had to be informed before the presentation of each screen where the fixation point would be for that screen. Moreover, this design does not damage the reliability or validity of the results.

The computer programs handled all matters of stimulus screen presentation and recording of observers' responses, which were acquired using the number keys on the keyboard. Each observer was allowed to view the stimulus screens at his or her own pace. To start a test sitting, one had to simply press Enter. Then, after a half-second delay, a stimulus screen would be presented for 200 ms. After another half-second delay, the observer would be prompted for his or her guess regarding the number of dots in the box for the screen just presented. After typing this number on the keyboard, the observer would press Enter again to repeat the process, until 300 stimulus screens had been presented. One sitting of 300 screens took approximately 30 minutes to complete.

The following directions were given to each observer when he or she was about to begin his or her first test sitting, which was always task 1:

> You are going to see some dots presented very quickly on the computer screen; they will appear in a rectangular box near the bottom of the screen. Look directly at the center of the box. For each presentation of dots, try to count the number of dots you see. If you can't get an exact number, just take your best guess. Type this number into the computer when it asks you to, and then press Enter to see the next group of dots. To help you count the dots, remember this: there will always be at least one dot in the box, and there will never be more than 30. So your guesses should range only from 1 to 30. When you're ready to begin, press Enter.

Later, when an observer was about to start a task 2 & 3 sitting, the following further directions were given:

> Now you're going to see some more groups of dots, but there are a few changes this time. Look at the middle of the very top of the screen, where there will be a plus sign, even though the box with the dots will appear at the bottom as usual. Make sure you look only at the plus sign, *not* the box. Also, sometimes, dots will appear around the box in addition to within it. Your task is still to *count only the dots in the box*. So when there are dots around the box, try to ignore them. Again, if you can't get an exact number for a given group of dots in the box, just guess. Other than that, your task is the same, and the range of dot numbers will still be 1 to 30. Enter your responses just as you did before.

One might ask why we told our observers the domain of numerosities beforehand. We assumed that if we did not do this, each observer would make his or her own independent guess as to the maximum number of dots shown in the box. These guesses would vary between the observers, possibly greatly, which would destabilize the results. Telling observers the maximum number of dots gave them a frame of reference from which to judge the number in each presentation.

Results

Compared to counting uncluttered objects in central vision (task 1), we find that we are somewhat worse at counting uncluttered objects in peripheral vision (task 2), and a lot worse at counting cluttered objects in peripheral vision (task 3).

The following analysis was employed to arrive at these results. At each numerosity of each task, each observer's error was calculated, and the mean of the five observers' errors was calculated from these. (Recall that *error* is the sample standard deviation of an observer's 30 counts at each numerosity of each task.) Figure 3 panels (a), (b), and (c) show this step for tasks 1, 2, and 3, respectively. The calculated mean error values from all three tasks were then plotted on a single pair of axes, and lines of best fit were drawn. This is figure 4.

There are several interesting aspects of these data. First, counting objects out of peripheral vision instead of central vision yields an uneven impairment. This is represented graphically by the uneven upward shift of the task 2 error curve from the task 1 error curve in figure 4. For numerosities above three, error is constantly greater by about one unit, but for numerosities 1, 2, and

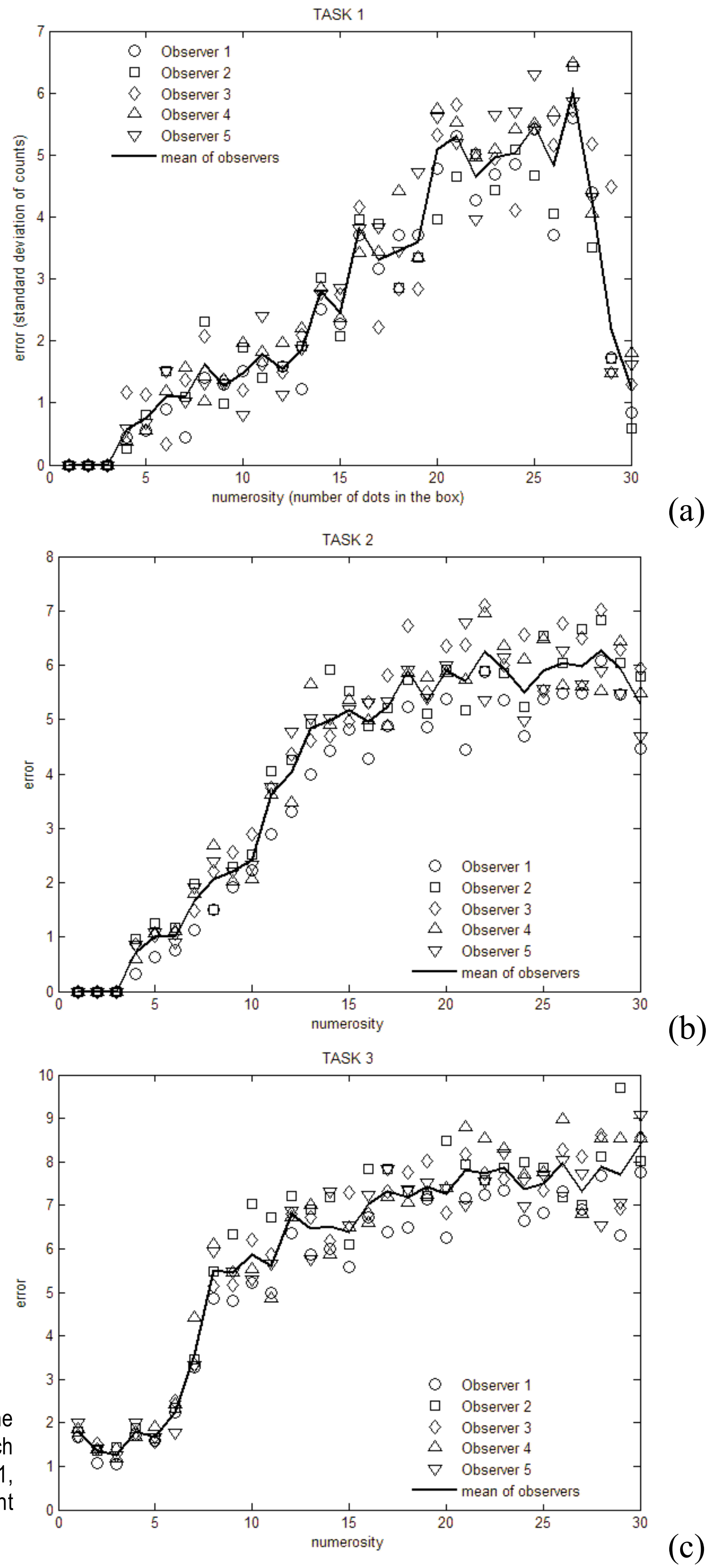

Figure 3. Each observer's error data and the five observers' mean error for each numerosity for the three tasks; (a) is task 1, (b) is task 2, (c) is task 3. Note the different scales on the graphs' vertical axes.

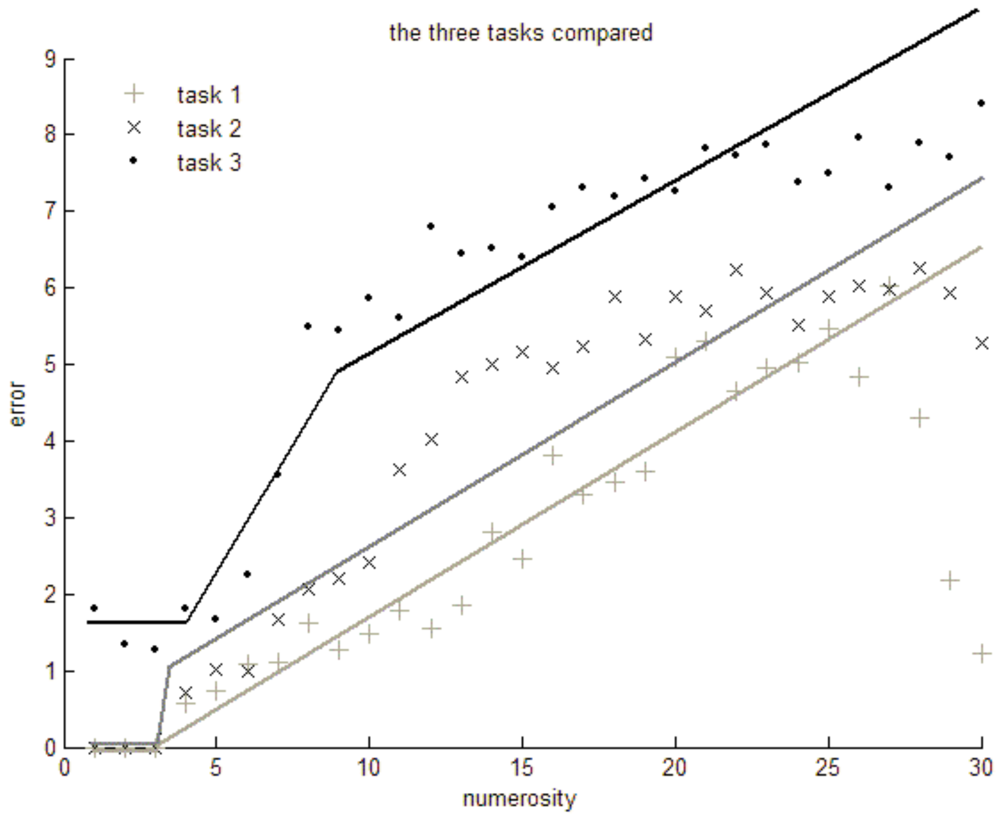

Figure 4. Lines of best fit for the five observers' mean error on each task. The data points from which the lines are drawn, seen on the graph as plus signs, x's, and small black dots, match the vertices of the thick black "mean error" lines in figure 3.

Tasks 1 and 2 both have zero error over the domain of numerosities from 1 through 3. However, for clarity, the lines of best fit are slightly offset from each other there.

3, the errors are identical: zero. Second, a similar uneven impairment is observed between counting cluttered versus uncluttered objects out of peripheral vision, represented by the uneven upward shift of the task 3 error curve from the task 2 error curve. This brings up two interesting points. For numerosities above about 8, the difference in error between task 2 performance and task 3 performance is about two and a half units, much greater than the difference between task 1 performance and task 2 performance. But the difference between the two tasks over the numerosities up to 3 is even more striking: in task 3, for the first time, we observe non-zero error in this domain. The observers were making errors counting groups of 1, 2, and 3 dots. This will be addressed further in the discussion.

Finally, observe that for all three tasks, in the domain of numerosities above about 25, the slopes of the mean error lines decrease (not the best fit lines, but the dark black lines shown in figure 3). In the case of task 1, the decrease is very great. We explain this as merely a mathematical consequence: because we told the observers that there would never be more than 30 dots in the box, they never gave responses greater than 30, even when they might have if they had not known the maximum number. This thinned the distributions of responses for numeosities close to 30, which artificially deflated the error values we obtained for these numerosities. This is why we considered this upper domain of numerosities little when we drew our best fit curves.

Statistical Analysis. For our statistical analysis, we performed one-sample *t*-tests with four degrees of freedom to test for statistical difference between the three best fit error curves. Specifically, we tested the statistical significance of the two impairments we observed: the difference between task 2 and task 1, and the difference between task 3 and task 2. Because nothing would be gained by doing this at every single numerosity, and because doing so would be time consuming and cumbersome, we choose three arbitrary numerosities at which to test the statistical significance: 10, 18, and 25. For example, consider the difference between tasks 2 and 3 at numerosity 18. Using the formula $t = (x - □) / (s / \sqrt{(n)})$, we substituted as follows: x = 7, the task 3 error at numerosity 18; □ = 4.5, the task 2 error at numerosity 18; s = 0.462738, the sample standard deviation of the individual observers' task 3 errors at numerosity 18; and n = 5, the number of observers. A computer

task 1 vs. task 2:

	x	□	s	*t*	*P*
10	2.6	1.7	0.317815	6.33	0.002
18	4.5	3.6	0.539771	3.73	0.0102
25	6.1	5.3	0.564533	3.17	0.017

Table 1. The values of the input and output variables in the *t*-tests that were used to validate statistically significant difference between the error curves of tasks 1 and 2. The difference is statistically significant, because all the *P*-values are less than the established alpha level of .05.

task 2 vs. task 3:

	x	□	s	*t*	*P*
10	5.2	2.6	0.761767	7.63	0.0008
18	7	4.5	0.462738	12.08	0.0001
25	8.4	6.1	0.418490	12.29	0.0001

Table 2. The values of the input and output variables in the *t*-tests that were used to validate statistically significant difference between the error curves of tasks 2 and 3. The difference is statistically significant, because all the *P*-values are less than the established alpha level of .05.

performed the actual calculations, and found *P*-values for each test as well. The alpha level of 0.05 was established for all tests. Tables 1 and 2 list the statistical test inputs and results. Note that every *P*-value obtained is well below the established alpha level, which indicates that the differences between all three error curves are statistically significant.

Discussion

We can now answer the questions with which we began. In general, counting uncluttered objects with peripheral vision impairs counting accuracy, but only slightly, adding about one to the standard deviation of the response distribution. However, for three objects and fewer, peripheral viewing does not affect accuracy; we still perform perfectly for such small numbers of objects. But this is not the case when counting cluttered objects peripherally. Counting in this condition is much more difficult than counting in either of the other conditions. For many objects, the standard deviation of one's response distribution is increased by more than two in comparison to uncluttered peripheral counting. (The task 3 curve is about two and a half units above the task two curve in figure 4.) Even more remarkably, this condition introduces error into counts of the smallest numbers of objects. One cannot reliably count even one, two, or three cluttered objects in a peripheral glance.

Indiscriminate Object Locations. There is some subjective evidence to suggest that viewing duration would not affect counting accuracy for cluttered objects in the periphery. Namely, regarding task 3, several observers reported that though they could clearly see the box and all the dots, they were strangely unable to determine which dots were in the box and which were outside it. One observer said, "The box and the dots are clearly there, I just can't tell whether the dots are in the box." Another added humorously, "The box is like a control-freak parent looking over your shoulder. You know it's always there, but never exactly where it is." If these observers were reporting accurately, the difficulty of the task may be due to an inability to locate individual objects in the periphery when the objects are cluttered. If this were the case, a longer viewing duration would probably not decrease counting error (if the observers did not move their fixation to the dots). This is because one wouldn't know which dots to mentally attend to—in this case, to count—if it were impossible to tell exactly where they were located. Investigators Intriligator and Cavanagh have studied this location ambiguity effect (2001).

Several investigators report that distracters are more effective at "masking" other objects, that is, rendering them less visually discriminable, the more similar they are to the objects be masked in terms of size, shape, and color (Nazir, 1992; Kooi *et al*, 1994). This property of our distracters (in task 3) probably had a lot to do with task 3's success in greatly impairing counting ability: the distracters were identical to the targets.

Separate Systems of Number Estimation? There has been some debate among psychophysicists regarding the mental systems that provide our ability to estimate number. (The word "systems" is used for lack of a better one, for the systems are essentially psychological constructs; little attempt is made to find their possible neurological bases.) One common view is that there are two distinct systems, *subitizing* and the *magnitude system*, each of which has its own properties and is effective only over a certain domain of object numerosities (see Feigenson, Dehaene, and Spelke, 2004, for a review). According to this theory, subitizing is the ability to instantaneously estimate a number of objects up to three or four—this is uncertain—with perfect accuracy. The magnitude system supports the ability to quickly obtain a reasonable, "order of magnitude" estimation for the quantity of a number of objects, but only when there are at least four objects. This is an elegant theory, and the systems seem to be separate based on findings from studies of infants' numerical ability (see, for example, Feigenson and Carey, 2003; Xu and Spelke, 2000).

But there is no proof that the systems are so dichotomous and well defined, or even that there are two distinct systems at all. Our results can be taken as evidence for this two-system theory.

Figure 5 is a graph of our results with error on a logarithmic scale. The data points and best fit curves are the same as those in figure 4; the only difference is the scaling of the vertical axis. With this visualization, it is clear that task 3's impairment is much greater in subitizing's domain (1 to 3) than in the magnitude system's domain (greater than 3). In fact, if we had considered a much larger domain of numerosities (say 1 to 1,000) and the trends we observe here proved consistent, then when graphing the results with a logarithmic error scale, the three tasks' error curves would be practically indistinguishable—except in the subitizing domain, where task 3's impairment would appear strong as ever. Therefore, there is a stark disparity in the magnitude of the impairment, and the rapid change in the impairment level occurs at the same numerosity at which subitizing no longer supports number estimation and the magnitude system begins to. This corroborates the other evidence that supports the two-system theory. A very logical explanation of our result is that something about task 3 did not allow the observers to subitize as well as usual—that is, it partially "knocked out" subitizing in their minds—but it impaired the magnitude system only slightly.

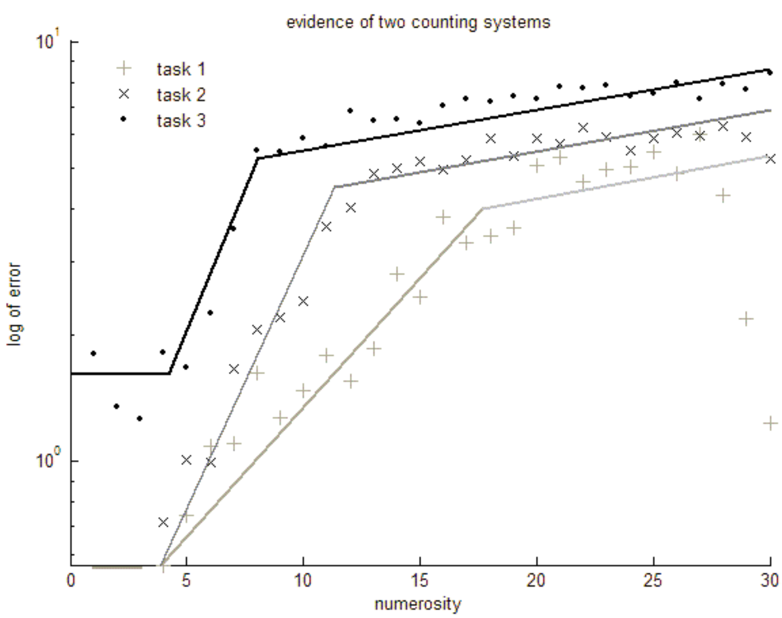

Figure 5. Lines of best fit for the five observers' mean error on each task, with error scaled logarithmically.

Conclusion. We measured accuracy in counting objects under various visual conditions. We find that human counting ability is very weak, much weaker than it appears in everyday use. When counting cluttered objects in the peripheral field of vision in a glimpse, there is not a single quantity of objects for which we perform anywhere near perfectly: there is great error even for counting a single object, and the error only increases with the number of objects to be counted. To ensure accuracy in counting a group of objects, central viewing and careful attention to each individual object, one at a time, are absolutely necessary.

Acknowledgements

To the following people I offer my warmest thanks:

- to Denis Pelli, for hosting me in his lab, for lengthily reviewing this paper, and for teaching me much about his field;
- to Kat Tillman, for her helpful comments at lab meetings;
- to Anne Manwell, for consistently impressing me with her consideration and concern;
- and to Steven Farbstein, for his refreshing pragmatism and utilitarian advice.

Thank you.

References

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision* 10:433-436.

Carpenter, R. H. S. (1988). *Movements of the Eyes*, London: Pion.

Feigenson, L., and Carey, S. (2003). Tracking Individuals via Object Files: evidence from infants' manual search. *Dev Sci* 6:568-584.

Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core Systems of Number. *Trends Cogn Sci* 8(7):307-314.

Intriligator, J., and Cavanagh, P. (2001). The Spatial Resolution of Visual Attention. *Cognit Psychol* 43(3): 171-216.

Jevons, W. S. (1871). The Power of Numerical Discrimination. *Nature* 3:281-282.

Kooi, F. L., Toet, A., Tripathy, S. P., and Levi, D. M. (1994). The Effect of Similarity and Duration on Spatial Interaction in Peripheral Vision. *Spatial Vision* 8:255–279.

Nazir, T. A. (1992). Effects of Lateral Masking and Spatial Precueing on Gap-resolution in Central and Peripheral Vision. *Vision Research* 32:771–777.

Pelli, D. G. (1997). The VideoToolbox Software for Visual Psychophysics: Transforming Numbers Into Movies. *Spatial Vision* 10:437-442.